

# Persistence of homoclinic orbits for billiards and twist maps

Sergey Bolotin<sup>†</sup>, Amadeu Delshams<sup>‡</sup> and Rafael Ramírez-Ros<sup>‡</sup>

<sup>†</sup> Department of Mathematics, University of Wisconsin-Madison, USA and  
Department of Mathematics and Mechanics, Moscow State University

<sup>‡</sup> Departament de Matemàtica Aplicada I, Universitat Politècnica de Catalunya,  
Diagonal 647, 08028 Barcelona, Spain

**Abstract.** We consider the billiard motion inside a  $C^2$ -small perturbation of a  $n$ -dimensional ellipsoid  $Q$  with a unique major axis. The diameter of the ellipsoid  $Q$  is a hyperbolic two-periodic trajectory whose stable and unstable invariant manifolds are doubled, so that there is a  $n$ -dimensional invariant set  $W$  of homoclinic orbits for the unperturbed billiard map. The set  $W$  is a stratified set with a complicated structure.

For the perturbed billiard map the set  $W$  generically breaks down into isolated homoclinic orbits. We provide lower bounds for the number of primary homoclinic orbits of the perturbed billiard which are close to unperturbed homoclinic orbits in certain strata of  $W$ .

The lower bound for the number of persisting primary homoclinic billiard orbits is deduced from a more general lower bound for exact perturbations of twist maps possessing a manifold of homoclinic orbits.

AMS classification scheme numbers: 37J15, 37J40, 37J45, 70H09

PACS numbers: 05.45.-a, 45.20.Jj, 45.50.Tn

E-mail: bolotin@math.wisc.edu, Amadeu.Delshams@upc.es, rafael@vilma.upc.es

## 1. Introduction

Billiards are commonly considered as one of the most standard frameworks to look for chaotic behavior. However, *elliptic billiards*—billiards inside  $n$ -dimensional ellipsoids—are by far the most famous example of discrete integrable systems. And it is also a very rare example since, according to *Birkhoff's conjecture*, among all billiards inside smooth convex hypersurfaces, only the elliptic ones are integrable [21, §2.4].

As it happens with all famous integrable systems, global action-angle variables cannot be introduced for elliptic billiards, due to the existence of several isolated invariant sets with some hyperbolic behavior. For instance, inside *ellipsoids with one diameter*—ellipsoids with a unique major axis—the diameter is a hyperbolic two-periodic billiard trajectory whose stable and unstable invariant manifolds are *doubled*, that is, they coincide as sets in the phase space forming a complicated stratified set.

Under a small perturbation of the ellipsoid, the hyperbolic two-periodic trajectory is only slightly shifted, but its invariant manifolds do not need to coincide, giving rise to the phenomenon called *splitting of separatrices*.

In the last years, several works [22, 16, 9, 15] were devoted to analyze the splitting of separatrices for planar billiards inside perturbed ellipses. In particular, in [9] (see also [8]) it was obtained a local version of the Birkhoff's conjecture. Concretely, it was shown that the billiard motion inside the perturbed ellipse

$$x = a \cos \phi \quad y = b(1 + \epsilon \eta(\phi)) \sin \phi$$

becomes non-integrable for any non-constant entire  $\pi$ -periodic function  $\eta : \mathbb{R} \rightarrow \mathbb{R}$ .

The study of the splitting of separatrices for higher-dimensional billiards was initiated in [11], which was focused on billiards inside perturbations of *prolate ellipsoids*, that is, ellipsoids with all the axis equal except for the major one. Prolate elliptic billiards are the simplest higher-dimensional generalization of planar elliptic billiards, because, on account of the conservation of the angular momenta, they are very similar to the planar ones. Later on, some results on *generic ellipsoids*—ellipsoids without equal axis—were obtained in [8], although only for symmetric perturbations.

The basic tool of those works is a twist discrete version of the Poincaré-Melnikov-Arnold method, which provides a *Melnikov potential*, whose non-degenerate critical points give rise to transverse homoclinic orbits to the diameter. The computation of the critical points of the Melnikov potential is feasible for perturbed ellipses and some symmetric perturbed prolate ellipsoids [11, 8], but becomes very intricate for general perturbed ellipsoids.

The main objective of this paper is to study the disintegration of the homoclinic set  $W$  and to provide a general lower bound for the number of primary homoclinic billiard orbits to the diameter that persist in billiard maps inside perturbed ellipsoids. This bound holds for any  $C^2$  perturbation and does not need any first order computation in the perturbation parameter like in the Melnikov method. This lower bound is obtained by means of a variational approach and Ljusternik-Schnirelmann theory.

We next describe more precisely the results of our paper, which were announced in [5]. Let

$$Q = \left\{ q = (q_0, \dots, q_n) \in \mathbb{R}^{n+1} : \sum_{i=0}^n \frac{q_i^2}{d_i^2} = 1 \right\} \quad (1)$$

be the  $n$ -dimensional ellipsoid which is supposed to *have one diameter*:

$$d_0 > d_1 \geq \dots \geq d_n > 0. \quad (2)$$

Denote  $\ell(Q) = 4s + 3m$ , where  $s$  is the number of the single axis among  $d_1, \dots, d_n$ , and  $m$  is the number of the multiple ones counted without multiplicity.

**Theorem 1.** *Inside any  $C^2$ -small perturbation of the ellipsoid  $Q$  there exist at least  $\ell(Q)$  primary homoclinic billiard orbits.*

The number  $\ell(Q)$  runs from 3 to  $4n$ . In the generic case

$$d_0 > d_1 > \dots > d_n > 0,$$

$s = n$ ,  $m = 0$ , so that  $\ell(Q) = 4n$ , whereas in the prolate case

$$d_0 > d_1 = \dots = d_n > 0,$$

$s = 0$ ,  $m = 1$ , so that  $\ell(Q) = 3$ .

We remark that all homoclinic orbits found in this paper are *primary*, that is, they exist for all the values of the perturbation parameter  $\epsilon$ , once assumed small enough, and they tend, along a subsequence, to unperturbed homoclinic orbits as  $\epsilon \rightarrow 0$ . It is well-known that the existence of transverse primary homoclinic orbits implies the existence of an infinite number of multi-bump homoclinic orbits. These homoclinic orbits usually do not have a limit as  $\epsilon \rightarrow 0$  and their dependence on  $\epsilon$  is more complicated. The bifurcation of such *secondary homoclinic* orbits will be described in another paper.

Our results are deduced from a general theorem on the persistence of heteroclinic orbits for exact perturbations of twist maps. Let  $f : M \rightarrow M$  be a twist map with a couple of hyperbolic periodic orbits whose invariant manifolds have a clean intersection along an invariant set  $N$ . We recall that two submanifolds  $L_1$  and  $L_2$  of  $M$  have a *clean* intersection along  $N \subset L_1 \cap L_2$  if and only if each connected component of  $N$  is a submanifold of  $M$  and  $T_x N = T_x L_1 \cap T_x L_2$  for any  $x \in N$ . If  $N$  verifies certain compactness hypotheses, then for any  $C^1$ -small exact perturbation of  $f$  there exist at least  $\text{cat}(N/f)$  primary heteroclinic orbits tending to  $N$  as  $\epsilon \rightarrow 0$ . Here  $\text{cat}(N/f)$  is the Ljusternik-Schnirelmann category of the quotient space.

This result generalizes a previous one by Xia [24] for general symplectic maps, which holds only when the unperturbed invariant manifolds are *completely doubled*. (See section 3 for this definition. In the case of billiard maps on ellipsoids, the invariant manifolds are completely doubled only in the prolate case.)

Heteroclinic orbits of twist maps are critical points of the action functional on the Hilbert manifold of bi-infinite sequences in the configuration space satisfying certain asymptotic behavior at their ends. Using that the unperturbed invariant manifolds

have a clean intersection along  $N$ , it can be deduced that the unperturbed action has several finite-dimensional *non-degenerate critical manifolds* in the sense of Bott [6]. The compactness hypotheses are used to check that the quotient set  $K$  of these critical manifolds under translation is a union of compact manifolds. Thus, for the perturbed twist map, the primary heteroclinic orbits close to  $N$  correspond to critical points of a function (called splitting potential) defined on  $K$ . The Ljusternik-Schnirelmann category of  $K$  is the bound  $\text{cat}(N/f)$  we were looking for.

The first order term of the splitting potential is the Poincaré-Melnikov potential which can be often explicitly computed. Its non-degenerate critical points give non-degenerate critical points of the splitting potential, and so, transverse primary heteroclinic orbits. In the homoclinic case, according to the *Birkhoff-Smale theorem* [20], the map is *chaotic*, that is, its restriction to some invariant Cantor set is conjugate to a transitive topological Markov chain. However, recent results [7] show that instead of transversality, the existence of topological crossings between the stable and unstable invariant manifolds is enough for the existence of chaotic motions. From such kind of results, it appears that billiards inside any *entire* non-quadratic perturbation of ellipsoids with one diameter are chaotic (see [8] for related results and techniques).

Our results hold for any  $C^1$ -small exact symplectic perturbation of the billiard map  $f$ . Suppose for example that the motion of a point in the interior  $D_\epsilon$  of the surface  $Q_\epsilon$  is governed by a Hamiltonian system  $(H_\epsilon)$  with Hamiltonian  $H_\epsilon(q, p)$  which is a  $C^2$ -small perturbation of the free motion Hamiltonian:  $H_0(q, p) = |p|^2/2$ . Then for fixed energy  $E > 0$  the billiard motion inside  $B_\epsilon$  is governed by an exact symplectic map  $f_\epsilon$  of  $T^*Q_\epsilon$  with generating function  $L_\epsilon$  close to  $L$ . The generating function  $L_\epsilon$  is defined as follows. For two points  $q_1, q_2 \in Q_\epsilon$  take the solution  $\gamma(t) = (q(t), p(t))$ ,  $t_1 \leq t \leq t_2$ , of system  $(H_\epsilon)$  with energy  $E$  joining these points:  $q(t_1) = q_1$ ,  $q(t_2) = q_2$ . Then

$$L_\epsilon(q_1, q_2) = \int_\gamma p dq$$

is the Maupertuis action of  $\gamma$ . For  $\epsilon = 0$  we get  $L_0(q_1, q_2) = \sqrt{2E}|q_1 - q_2|$ .

Suppose for example that a charged particle moves in  $\mathbb{R}^3$  under the influence of a small stationary electro-magnetic field. Then the magnetic field has a vector potential  $A_\epsilon$ , and the electric field has a scalar potential  $V_\epsilon$ . Thus the motion of the particle is governed by the Hamiltonian equations with Hamiltonian

$$H_\epsilon(q, p) = \frac{1}{2}|p - A_\epsilon(q)|^2 + V_\epsilon(q)$$

and our results imply the following corollary.

**Corollary 1.** *Under the influence of any small enough electro-magnetic field and inside any  $C^2$ -small perturbation of an ellipsoid in  $\mathbb{R}^3$  with a unique major axis, there exist at least 3—8 for generic ellipsoids—primary homoclinic billiard orbits.*

The same holds if the ellipsoid is slowly rotating around a fixed axis with small angular velocity. Indeed, in a rotating coordinate frame the Coriolis force is a magnetic force. Planar billiards in constant magnetic fields were first considered in [19], although

only inside planar regions. The splitting of separatrices in a slowly rotating planar ellipse was studied in [14].

We finish this introduction with the organization of the paper. In section 2, we state the results about the persistence of homoclinic orbits for billiards, whose proofs have been relegated to section 5. In section 3, we present the results on the persistence of heteroclinic orbits for twist maps. The proof of the main theorem of that section is contained in section 4.

## 2. Persistence of homoclinic orbits for billiards

Our results on billiards are described here. Firstly, we shall introduce convex billiards in a standard way. Secondly, we shall recall the main properties of elliptic billiards we are interested in. Finally, we shall give the lower bound on the number of persistent primary homoclinic orbits under  $C^2$ -small perturbations of ellipsoids with one diameter.

### 2.1. Convex billiards

Let  $Q$  be a  $C^2$  closed convex hypersurface in  $\mathbb{R}^{n+1}$ . Consider a particle moving freely inside  $Q$  and colliding elastically with  $Q$ , that is, at the impact points the velocity is reflected so that its tangential component remains the same, while the sign of its normal component is changed. This motion can be modelled by means of a  $C^1$  diffeomorphism  $f : M \rightarrow M$  defined on the  $2n$ -dimensional phase space

$$M = \{m = (q, p) \in Q \times \mathbb{S}^n : p \text{ is directed outward } Q \text{ at } q\} \quad (3)$$

consisting of impact points  $q \in Q$  and unitary velocities  $p \in \mathbb{S}^n$ .

The *billiard map*  $f(q, p) = (q', p')$  is defined as follows. The new velocity  $p' \in \mathbb{S}^n$  is the reflection of  $p \in \mathbb{S}^n$  with respect to the tangent plane  $T_q Q$ , and the new impact point  $q' \in Q$  is determined by imposing  $p' = (q' - q)/|q' - q|$ . The existence and uniqueness of the point  $q'$  follows from the convexity and closeness of the hypersurface  $Q$ . The map  $f$  is symplectic:  $f^* \omega = \omega$ , where  $\omega$  is the symplectic form  $\omega = d\alpha$  and  $\alpha = p \cdot dq$ .

Two consecutive impact points  $q$  and  $q'$  determine uniquely the velocity  $p'$ , and hence the following impact point  $q''$ . Thus, one can also define the billiard map in terms of couples of consecutive impacts points by  $f : (q, q') \mapsto (q', q'')$  on the phase space  $U = Q^2 \setminus \Delta$ , where  $\Delta = \{(q, q') \in Q^2 : q = q'\}$ .

A *billiard orbit* is a bi-infinite sequence  $(m_k) \in M^{\mathbb{Z}}$  such that  $f(m_k) = m_{k+1}$ . A *billiard trajectory* is a bi-infinite sequence of impact points  $(q_k) \in Q^{\mathbb{Z}}$  such that  $f(q_k, p_k) = (q_{k+1}, p_{k+1})$  for  $p_{k+1} = (q_{k+1} - q_k)/|q_{k+1} - q_k|$ . Billiard orbits and billiards trajectories are in one-to-one correspondence, so we can use them indistinctly.

The *chords* of the hypersurface  $Q$  are the segments perpendicular to  $Q$  at their ends. The longest chords are called *diameters*. Any chord gives rise to a two-periodic set. If the chord is a *diameter*, the two-periodic set is usually hyperbolic.

## 2.2. Elliptic billiards

Let  $f : M \rightarrow M$  be the billiard map inside the  $n$ -dimensional ellipsoid (1) with one diameter. Introducing the matrix  $D = \text{diag}(d_0, \dots, d_n)$ , the ellipsoid can be expressed by the equation  $\langle q, D^{-2}q \rangle = 1$ . It is useful to group together the eigenvalues of  $D$ . We write  $D = \text{diag}(\tilde{d}_0, \tilde{d}_1 \cdot \text{Id}_{s_1}, \dots, \tilde{d}_l \cdot \text{Id}_{s_l})$ , where

$$d_0 = \tilde{d}_0 > \tilde{d}_1 > \dots > \tilde{d}_l > 0 \quad s_1 + \dots + s_l = n$$

and  $\text{Id}_r$  stands for the  $r \times r$  identity matrix. Hence  $s_j$  is the multiplicity of the eigenvalue  $\tilde{d}_j$ . Denote  $s(Q) = (1, s_1, \dots, s_l) \in \mathbb{N}^{l+1}$ . We also introduce the following  $l$  couples of natural numbers

$$\left. \begin{aligned} a_j &= 1 + s_1 + \dots + s_{j-1} \\ b_j &= s_1 + \dots + s_{j-1} + s_j \end{aligned} \right\} \quad j = 1, \dots, l. \quad (4)$$

These couples are determined by the conditions

$$d_i = \tilde{d}_j \iff i \in \llbracket a_j, b_j \rrbracket := \mathbb{Z} \cap [a_j, b_j] = \{a_j, \dots, b_j\}. \quad (5)$$

We note that  $s_j = \#\{i : d_i = \tilde{d}_j\} = \#\llbracket a_j, b_j \rrbracket = b_j - a_j + 1$ .

The least degenerate ellipsoids with one diameter are the generic ellipsoids:

$$d_0 > d_1 > \dots > d_n > 0 \implies s(Q) = (1, \dots, 1) \in \mathbb{N}^{n+1}$$

whereas the most degenerate ones are the prolate ellipsoids:

$$d_0 > d_1 = \dots = d_n > 0 \implies s(Q) = (1, n) \in \mathbb{N}^2.$$

The two-periodic orbit associated to the chord joining the vertices  $(-d_0, 0, \dots, 0)$  and  $(d_0, 0, \dots, 0)$  is  $P = \{m_+, m_-\}$ , where  $m_\pm = (q_\pm, p_\pm)$ ,  $q_\pm = (\pm d_0, 0, \dots, 0)$ , and  $p_\pm = (\pm 1, 0, \dots, 0)$ . Obviously,  $f(m_\pm) = m_\mp$ . The spectrum of  $Df^2(m_\pm)$  has the form

$$\{\lambda_1^2, \dots, \lambda_n^2, \lambda_1^{-2}, \dots, \lambda_n^{-2}\}$$

where  $\lambda_1 \geq \dots \geq \lambda_n > 1$  are the *characteristic multipliers* of  $P$ , namely

$$\lambda_i = (1 + e_i)(1 - e_i)^{-1} \quad e_i = (1 - d_i^2/d_0^2)^{1/2} \quad (6)$$

Thus, the periodic orbit  $P$  is *hyperbolic* if and only if  $d_0 > d_i$  for  $i = 1, \dots, n$ , or equivalently, if and only if the chord joining  $(-d_0, 0, \dots, 0)$  and  $(d_0, 0, \dots, 0)$  is the unique diameter of the ellipsoid. For this reason, the problem of splitting of separatrices we shall deal with is well-posed only for ellipsoids with one diameter.

Elliptic billiards are completely integrable. Such integrability is closely related to a property of confocal quadrics, see [21, §2.3]. We only need the following well-known family of first integrals

$$F_i(m) = p_i^2 + \sum_{i' \neq i} \frac{(q_i p_{i'} - q_{i'} p_i)^2}{d_i^2 - d_{i'}^2} \quad i = 0, \dots, n \quad (7)$$

where  $m = (q, p)$ ,  $q = (q_0, \dots, q_n)$ , and  $p = (p_0, \dots, p_n)$ .

These first integrals are dependent:  $\sum_{i=0}^n F_i(m) = \sum_{i=0}^n p_i^2 = |p|^2 \equiv 1$ , but skipping one of them the rest are independent almost everywhere. Unfortunately, they are well-defined only for generic ellipsoids, on account of the presence of the denominators  $d_i^2 - d_{i'}^2$ .

Due to that, degenerate ellipsoids are often avoided in the literature. Nevertheless, in the presence of degenerations, it suffices to substitute the first integrals that become singular by some regular ones.

To be more precise, if  $s_j = \#\llbracket a_j, b_j \rrbracket > 1$ , then we substitute the singular integrals  $F_i$ ,  $i \in \llbracket a_j, b_j \rrbracket$ , by their regular sum

$$S_j(m) = \sum_{i \in \llbracket a_j, b_j \rrbracket} F_i(m) = \sum_{i \in \llbracket a_j, b_j \rrbracket} \left( p_i^2 + \sum_{i' \notin \llbracket a_j, b_j \rrbracket} \frac{(q_i p_{i'} - q_{i'} p_i)^2}{\tilde{d}_j^2 - d_{i'}^2} \right)$$

and the angular momenta

$$K_{(i,i')}(m) = q_i p_{i'} - q_{i'} p_i \quad i, i' \in \llbracket a_j, b_j \rrbracket \quad i \neq i'. \quad (8)$$

From now on,  $A_j$  is the set where all the angular momenta (8) vanish and

$$Z_j = \begin{cases} F_i^{-1}(0) & \text{if } s_j = 1 \text{ and } i = a_j = b_j \\ S_j^{-1}(0) \cap A_j & \text{if } s_j > 1 \end{cases} \quad (9)$$

Among the function  $S_j$  and the angular momenta (8) there are  $s_j$  functions independent almost-everywhere, so the set  $Z_j$  has dimension  $2n - s_j$ . The hyperbolic periodic orbit  $P$  is contained in the  $n$ -dimensional level set

$$Z = \cap_{j=1}^l Z_j. \quad (10)$$

This level set plays an important role in the description of the  $n$ -dimensional *unstable and stable invariant manifolds*

$$W^- = W_f^u(P) = \left\{ m \in M : \lim_{k \rightarrow -\infty} \text{dist}(f^k(m), P) = 0 \right\}$$

$$W^+ = W_f^s(P) = \left\{ m \in M : \lim_{k \rightarrow +\infty} \text{dist}(f^k(m), P) = 0 \right\}$$

and the *homoclinic set*  $H = (W^- \cap W^+) \setminus P$ .

From the above definitions, it is clear that  $H \cup P = W^+ \cup W^- \subset Z$ . What is not so obvious is that these inclusions are, in fact, equalities:

$$H \cup P = W^- = W^+ = Z.$$

(This result is stated without proof, since it will not be used.) Following the classic terminology, we say that  $W^-$  and  $W^+$  are *doubled*.

The set  $H \cup P$  is an  $n$ -dimensional stratified set. It is not necessary to describe it in detail (see [8]), since in this paper we study a subset  $N$  of  $H$  given by

$$N = \cup_{j=1}^l N_j \quad N_j = H \cap \Pi_j = (Z_j \setminus P) \cap \Pi_j \quad (11)$$

where

$$\Pi_j = \{m \in M : q_i = p_i = 0 \text{ for all } i \notin \{0\} \cup \llbracket a_j, b_j \rrbracket\}. \quad (12)$$

The set  $N_j$  has a simple dynamical meaning. Let us consider the coordinate sections of the form

$$Q_j = \{q \in Q : q_i = 0 \text{ for all } i \notin \{0\} \cup \llbracket a_j, b_j \rrbracket\} \quad j = 1, \dots, l.$$

If two consecutive impact points are on  $Q_j$ , the same happens to all the impact points. Therefore, the set (12) is invariant by the map  $f$ , and  $f|_{\Pi_j}$  is a billiard map inside a  $s_j$ -dimensional ellipsoid. If  $s_j = 1$ , the corresponding sub-billiard is a planar one, and if  $s_j > 1$  it is prolate. Thus  $N_j$  is the homoclinic set of  $f|_{\Pi_j}$ , and  $N = \cup_{j=1}^l N_j$  is the union of the homoclinic sets of the (planar or prolate) sub-billiards associated to the partition  $\{1, \dots, n\} = \cup_{j=1}^l \llbracket a_j, b_j \rrbracket$ .

**Lemma 1.** *The set  $N = \cup_{j=1}^l N_j$  verifies the following properties:*

- (i)  $N_j$  is a  $s_j$ -dimensional submanifold of  $M$  invariant by  $f$ .
- (ii)  $N \cup P$  is compact.
- (iii) Given any neighbourhood  $V$  of  $P$ , there exists  $k_0 > 0$  such that  $f^k(N \setminus V) \subset V$  for all integer  $|k| > k_0$ .

These properties are local, that is, they only give local information about  $N$  and the action of the map  $f$  on it. The next property is more global, in the sense that it describes how the invariant manifolds  $W^-$  and  $W^+$  intersect along  $N$ .

**Lemma 2.** *The invariant manifolds  $W^-$  and  $W^+$  have a clean intersection along  $N$ .*

The proof of these lemmas are contained in subsections 5.4 and 5.5, respectively. These lemmas play a fundamental rôle in the proof of the main theorem on billiards stated below. Indeed, we will show that the persistence result holds for any twist map verifying these hypothesis, see theorem 4 and the remarks following it.

### 2.3. The theorem

Once we know that there is a  $n$ -dimensional set of homoclinic billiard orbits inside an ellipsoid with one diameter, it is quite natural to ask if some of those orbits persist under small perturbations of the ellipsoid.

We say that  $Q_\epsilon$  is a  $C^2$ -small perturbation of the ellipsoid  $Q$  when it is a  $C^2$  hypersurface of  $\mathbb{R}^{n+1}$  which is  $O(\epsilon)$ -close to  $Q$  in the  $C^2$ -topology. Then  $Q_\epsilon$  is also convex, has a unique diameter, and its corresponding billiard map is well-defined. In the following theorem, which is a slightly more precise version of Theorem 1, we give a lower bound on the number of primary homoclinic billiard orbits to the diameter.

**Theorem 2.** *Let  $1, s_1, \dots, s_l$  be the multiplicities of the axis of the ellipsoid  $Q$ . Then for any  $C^2$ -small perturbation of  $Q$  there are at least*

$$\ell_j = \begin{cases} 4 & \text{if } s_j = 1 \\ 3 & \text{if } s_j > 1 \end{cases}$$

*primary homoclinic billiard orbits close to  $N_j$ .*

The proof is relegated to subsection 5.1, since it involves some results on twists maps we have not explained yet.



### 3. Persistence of heteroclinic orbits for twist maps

In this section we prove two general results on perturbations of manifolds of homoclinic orbits for twist maps. These results hold for general exact symplectic maps. However, for simplicity we consider only twist maps with globally defined generation functions. Our choice of generality is motivated by applications to perturbations of billiards in ellipsoids.

#### 3.1. Twist maps

Let  $Q$  be a smooth  $n$ -dimensional manifold and  $\tilde{Q}$  its universal covering with the group  $G$  of covering transformations. Let  $U$  be an open set in the quotient space  $(\tilde{Q} \times \tilde{Q})/G$ , and let  $L \in C^2(U)$ . It can be viewed as a function on  $\tilde{Q} \times \tilde{Q}$  invariant under the group action:  $L(q, q') = L(g(q), g(q'))$  for any  $g \in G$ .

The *twist map* with the generating function  $L$  is the symplectic map  $f : M \subset T^*Q \rightarrow T^*Q$  defined as follows: for  $(q, p) \in T^*Q$  we set  $f(q, p) = (q', p')$  if

$$p = -D_1 L(q, q'), \quad p' = D_2 L(q, q'). \quad (13)$$

The map  $f$  is correctly defined on  $M$  if, for all  $(q, p) \in M$ , equations (13) have a unique solution  $(q', p') \in T^*Q$ . Equation (13) can be solved locally provided that the *local twist condition* holds: the Hessian

$$D_{12} L(q, q') \in \text{Hom}(T_q Q, T_{q'}^* Q)$$

is non-degenerate for all  $(q, q')$ . In general equations (13) can have several solutions, and the corresponding twist map  $f$  is multi-valued.

Let  $U_q = \{q' \in \tilde{Q} : (q, q') \in U\}$ . The twist map is a correctly defined single valued map  $f : T^*Q \rightarrow T^*Q$  if  $L$  satisfies the *global twist condition in  $U$* : for all  $q \in \tilde{Q}$  the map  $f_q : U_q \rightarrow T_q^* Q$  given by  $q' \mapsto D_1 L(q, q')$  is a diffeomorphism. The global twist condition can hold only if  $U_q$  is diffeomorphic to  $\mathbb{R}^n$ . In general the twist map  $f$  is well-defined on an open of  $T^*Q$ .

*Example 1.* For the  $n$ -dimensional billiard in a convex hypersurface  $Q$ , the generating function  $L$  satisfies the local twist condition in the open set  $U = (Q \times Q) \setminus \Delta$ . Since  $Q \simeq \mathbb{S}^n$ , the global twist condition cannot hold. However, the map  $f_q$  is a diffeomorphism of  $U_q = Q \setminus \{q\}$  onto the set  $\{p \in T_q^* Q : |p| < 1\}$ . Hence  $f$  is a symplectic diffeomorphism defined on an open  $T^*Q$  which is isomorphic to the phase space (3). For the billiard,  $L(q, q') = |q - q'|$  is a single valued function on  $Q \times Q$ , so there is no need to pass to a covering  $\tilde{Q}$ .

There is another definition of twist map which will be used in this section. Define the map  $g : U \rightarrow U$  by the formula

$$g(q, q') = (q', q''),$$

where  $q'' \in \tilde{Q}$  is determined by the equation

$$D_2 L(q, q') + D_1 L(q', q'') = 0. \quad (14)$$

If the global twist condition holds on  $U$ , then the map  $h : U \rightarrow M$  given by  $h(q, q') = (q, -D_1 L(q, q'))$  is a diffeomorphism, and  $h$  is a conjugacy between  $f$  and  $g$ . The map  $g$  is also called the twist map defined by the generating function  $L$ . We identify  $U$  and  $M$  via  $h$  and do not make any difference between  $f$  and  $g$ .

An orbit of the twist map  $f : M \rightarrow T^*Q$  is a sequence  $O = (m_i = (q_i, p_i) \in M)_{i \in \mathbb{Z}}$  such that  $m_i = f(m_{i-1})$ . The corresponding sequence  $q = (q_i \in \tilde{Q})$  will be called a trajectory. Even when the generating function  $L \in C^2(U)$  does not satisfy the twist condition, so that the twist map  $f$  is not well-defined on  $M$ , we can define its trajectory as a sequence  $q = (q_i \in \tilde{Q})_{i \in \mathbb{Z}}$  with  $(q_{i-1}, q_i) \in U$  and satisfying equation (14) at each step:

$$D_2 L(q_{i-1}, q_i) + D_1 L(q_i, q_{i+1}) = 0. \quad (15)$$

Formally, equation (15) means that  $q \in \tilde{Q}^{\mathbb{Z}}$  is a critical point of the action functional

$$S(q) = \sum_{i \in \mathbb{Z}} L(q_{i-1}, q_i).$$

Usually this series is divergent, and the functional does not make sense. However, there are many special situations when  $S$  does make sense. For example,  $S$  makes sense when we study periodic orbits. For a  $s$ -periodic sequence  $x = (x_i)$  we set

$$S(x) = \sum_{i=1}^s L(x_{i-1}, x_i).$$

Then  $x$  is an  $s$ -periodic trajectory if and only if  $S'(x) = 0$ . In the next section we represent homoclinic orbits of hyperbolic periodic orbits as critical points of  $S$ .

The semi-local results we discuss in this section hold also for multi-valued maps  $f : M \rightarrow T^*Q$  with a generating function  $L$  satisfying the local twist condition only near certain homoclinic trajectories. The global twist condition is never really used.

### 3.2. Perturbation of homoclinic orbits

We consider the existence of heteroclinic orbits to hyperbolic periodic orbits. Homoclinic orbits to hyperbolic periodic orbits is a particular case.

Suppose that the twist map  $f$  has two hyperbolic  $s$ -periodic orbits  $O^\pm = (m_i^\pm)$  with the same period:  $m_{i+s}^\pm = m_i^\pm$  for  $i = 1, \dots, s$ . They define periodic trajectories  $x^\pm = (x_i^\pm)$ ,  $x_{i+s}^\pm = x_i^\pm$ , satisfying (15). In applications to the billiard map,  $O^\pm$  are the diameter 2-periodic orbits and  $O^- = O^+$  up to a time shift:  $m_i^+ = m_{i+1}^-$ .

We assume that the local twist condition holds on  $O^\pm$ , i.e.  $\det D_{12} L(x_{i-1}^\pm, x_i^\pm) \neq 0$  for all  $i$ . Then the twist map  $f$  is well-defined in a neighborhood of each point  $m_i^\pm$ .

Suppose that the periodic orbits  $O^\pm$  are hyperbolic. Then every point  $m_i^\pm$  has  $n$ -dimensional stable and unstable manifolds  $W^{s,u}(m_i^\pm)$  in the phase space  $M$ . Denote  $W_i^- = W^u(m_i^-)$  and  $W_i^+ = W^s(m_i^+)$ . Then

$$W_i^\pm = \{m \in M : \text{dist}(f^k(m), f^k(m_i^\pm)) \rightarrow 0 \text{ as } k \rightarrow \pm\infty\}$$

and  $f(W_i^\pm) = W_{i\pm 1}^\pm$ . Hence  $f^s(W_i^\pm) = W_i^\pm$ . The stable and unstable manifolds of the periodic orbits  $O^\pm$  are  $W^\pm = \cup_{i=1}^s W_i^\pm$ . Fix some small  $\delta > 0$ . The local stable and unstable manifolds  $W_{i,\delta}^\pm$  consist of points whose positive or negative iterates respectively stay in a  $\delta$ -neighborhood of the periodic orbit. These manifolds are embedded disks in the phase space given by embeddings  $\phi_i^\pm : \mathbb{B}^n \rightarrow M$ , and  $f^{\pm 1}(W_{i,\delta}^\pm) \subset W_{i\pm 1,\delta}^\pm$  respectively. The global stable and unstable manifolds can be defined also as

$$W_i^- = \cup_{k \geq 0} f^{ks} W_{i,\delta}^-, \quad W_i^+ = \cup_{k \leq 0} f^{ks} W_{i,\delta}^+.$$

Hence  $W_i^\pm$  are immersed submanifolds in  $M$ , diffeomorphic to  $\mathbb{R}^n$ , and  $W_{i,\delta}^\pm$  is a  $\delta$ -ball in  $W_i^\pm$ . The topology on  $W_i^\pm$  is characterized as follows: a set  $V \subset W_i^\pm$  is open if and only if  $f^{\pm ks}(V) \cap W_{i,\delta}^\pm$  is open in  $W_{i,\delta}^\pm$  for all  $k > 0$ .

Suppose that the twist map has a heteroclinic orbit  $O = (m_i = (q_i, p_i))_{i \in \mathbb{Z}}$  from  $O^-$  to  $O^+$ . Then  $m_i \in W_i^+ \cap W_i^-$  and  $\text{dist}(m_i, m_i^\pm) \rightarrow 0$  as  $i \rightarrow \pm\infty$ . The heteroclinic orbit  $O$  defines the heteroclinic trajectory  $q = (q_i)_{i \in \mathbb{Z}}$  such that  $d(q_i, x_i^\pm) \rightarrow 0$  as  $i \rightarrow \pm\infty$  exponentially.

**Lemma 3.** *Suppose that the actions  $S(x_\pm)$  of the periodic orbits  $O^\pm$  coincide. Then without loss of generality we may assume that  $L(x_{i-1}^\pm, x_i^\pm) = 0$  for all  $i$ .*

*Proof.* Subtracting a constant from the generating function  $L$  we may assume that  $S(x_\pm) = 0$ . Then there exists a smooth function  $g$  on  $Q$  such that  $L(x_{i-1}^\pm, x_i^\pm) = g(x_i^\pm) - g(x_{i-1}^\pm)$ . Next we perform a calibration replacing the generating function  $L(q, q')$  by  $\tilde{L}(q, q') = L(q, q') + g(q) - g(q')$ . This does not change trajectories of the twist map in  $Q$  (orbits in  $T^*Q$  will change) but now  $\tilde{L}(x_{i-1}^\pm, x_i^\pm) = 0$  for all  $i$ .  $\square$

Without loss of generality, we will make this assumption in all this section. Then the *normalized action* of a heteroclinic trajectory  $q = (q_i)$  can be defined as

$$S(q) = \sum_{i \in \mathbb{Z}} L(q_{i-1}, q_i). \quad (16)$$

If the periodic orbits  $O^\pm$  are hyperbolic, then the series converges exponentially. The action  $S(q)$  is translation invariant: it doesn't change if  $q$  is replaced by its translation  $T(q) = (q_{i+1})$ .

Now we pass to the perturbation theory. We will assume that the unperturbed map  $f = f_0$  has a family of heteroclinic orbits. More precisely, suppose that  $f$  has an invariant manifold  $N = \cup_{i=1}^s N_i \subset W^+ \cap W^-$  with connected components  $N_i \subset W_i^+ \cap W_i^-$  and  $f(N_i) = N_{i+1}$ , where we set  $N_{s+i} = N_i$ . Then  $N$  consists of heteroclinic orbits from  $O^-$  to  $O^+$ . In applications,  $f$  is usually an integrable map, and  $N$  is contained in a critical level set of the first integrals.

Next consider a perturbation of the twist map  $f$ : a smooth exact symplectic map  $f_\epsilon$  that is  $C^1$ -close to  $f^\dagger$ . If  $\epsilon$  is small enough, then on any compact set  $V \subset U$ ,  $f_\epsilon$  is a twist map with the generating function

$$L_\epsilon = L + \epsilon L_1 + \epsilon^2 L_2, \quad (17)$$

$\dagger$  If  $M$  is non-compact, in order to define a distance on  $C^1(M, M)$ , we need to specify a Riemannian metric on  $M$ . We don't bother about this, since everything will happen in a compact subset of  $M$ .

with  $L_2$   $C^2$ -bounded as  $\epsilon \rightarrow 0$ . Hence  $L_\epsilon$  is  $C^2$ -close to  $L$ . By the implicit function theorem, for sufficiently small  $\epsilon$  the map  $f_\epsilon$  has hyperbolic  $s$ -periodic orbits  $O_\epsilon^\pm$  near  $O^\pm$ . We will assume that the actions of the perturbed periodic orbits coincide:  $S_\epsilon(O_\epsilon^-) = S_\epsilon(O_\epsilon^+)$ , where  $S_\epsilon$  is the action determined by the generating function  $L_\epsilon$ . Our goal is to prove the existence of *primary* heteroclinic orbits from  $O_\epsilon^-$  to  $O_\epsilon^+$ . Using lemma 3 we may assume without loss of generality that the generating function  $L_\epsilon$  is normalized, so that in particular  $L_1(x_{i-1}^\pm, x_i^\pm) = 0$ .

Our first result is a version of the Poincaré–Melnikov–Arnold theorem. We define the Poincaré–Melnikov potential  $P : N_i \rightarrow \mathbb{R}$  as

$$P(m) = \sum_{j \in \mathbb{Z}} L_1(q_{j-1}, q_j) \quad (18)$$

where  $f^j(m) = (q_j, p_j)$ . The next theorem belongs essentially to Poincaré.

**Theorem 3.** *If  $W^-$  and  $W^+$  have a clean intersection along  $N$ , then non-degenerate critical points of  $P$  correspond to transverse heteroclinic orbits of  $f_\epsilon$ . These orbits are smooth on  $\epsilon$ .*

In the degenerate case more hypotheses are needed.

**Theorem 4.** *Suppose that  $N$  satisfies the following conditions:*

- (i)  $W^-$  and  $W^+$  have a clean intersection along  $N$ .
- (ii)  $\overline{N} = N \cup O^- \cup O^+$  is compact in the topology of  $M$ .
- (iii) The twist condition holds on  $\overline{N}$ .
- (iv) The topologies in  $N$  induced from  $M$ ,  $W^+$ , and  $W^-$  coincide.

*Then the map  $f_\epsilon$  has at least  $\text{cat}(N/f)$  primary heteroclinic orbits from  $O_\epsilon^-$  to  $O_\epsilon^+$  close to  $N$ , for small enough  $\epsilon$ .*

*Remark 1.* Primary heteroclinic orbits for  $f_\epsilon$  are orbits which converge, along a subsequence, to heteroclinic orbits of  $f$  when  $\epsilon \rightarrow 0$ . Existence of secondary heteroclinic orbits needs some additional assumptions and will be discussed in another paper.

*Remark 2.* The quotient space  $N/f$  is the quotient under the  $\mathbb{Z}$ -action of the map  $f$  on  $N$ . Under the conditions of Theorem 4, the group action is discrete and  $N/f$  is a compact manifold diffeomorphic to  $N_i/f^s$ . Thus,  $\text{cat}(N/f) = \text{cat}(N_i/f^s)$ .

We shall call the last condition of the theorem the *finiteness condition*, since it can be reformulated as follows:

- (iv') *Given any neighborhood  $V$  of the periodic orbits  $O^-$  and  $O^+$ , there exists  $k_0 > 0$  such that  $f^k(N \setminus V) \subset V$  for all integer  $k$  with  $|k| \geq k_0$ .*

§  $P$  is also called the Melnikov potential. However, this function was introduced and widely used by Poincaré. The Melnikov function—derivative of  $P$ —, was used later by Melnikov. Poincaré was studying Hamiltonian systems, the case only slightly different from the case of symplectic maps.

This means that any heteroclinic orbit in  $N$  stays in the neighborhood  $V$ , except for a finite number of iterates. (Note that condition  $(iv')$  coincides with property  $(iii)$  of Lemma 1 in the section of billiards.) Of course, given any  $m \in N$  always there exists some  $k_0 = k_0(m) > 0$  such that  $f^k(m) \in V$  for all integer  $k$  with  $|k| \geq k_0$ . The finiteness condition means that  $k_0$  can be chosen independently of  $m \in N \setminus V$ . In other words, there exists some  $k_1 > 0$  such that  $N \subset f^{k_1}(V^-) \cup f^{-k_1}(V^+)$ , where  $V^\pm$  is the connected (in the  $W^\pm$ -topology) component of  $V \cap W^\pm$  that contains the periodic orbit  $O^\pm$ . To show the equivalence between the conditions  $(iv)$  and  $(iv')$ , we notice that if the three topologies coincide, then the compact set  $\overline{N} = N \cup O^- \cup O^+$  is covered by the union of open sets  $\cup_{k=0}^\infty (f^k(V^-) \cup f^{-k}(V^+))$ . Since there is a finite sub-covering, we get the finiteness condition.

It is very useful to find some cases in which the hypotheses of theorem 4 hold. A couple of simple cases is presented below.

As a first example, we consider heteroclinic orbits coming from unperturbed loops. A curve  $C \subset (W^- \setminus O^-) \cap (W^+ \setminus O^+)$  from a point in  $O^-$  to another point in  $O^+$  is a *non-degenerate loop* when

$$\dim(T_m W_- \cap T_m W_+) = 1 \quad \forall m \in C.$$

If  $C$  is non-degenerate loop,  $N = \cup_{i=0}^{s-1} f^i(C)$  verifies the hypotheses  $(i)$ - $(iv)$ . To check the finiteness condition, it suffices to realize that there exists a parameterization  $\gamma : \mathbb{R} \rightarrow C$  such that  $f^s(\gamma(t)) = \gamma(t+1)$  and  $\gamma(\pm\infty) \in O^\pm$ . Besides,  $N/f \simeq C/f^s \simeq \mathbb{R}/\mathbb{Z} \simeq \mathbb{S}^1$  and so the perturbed map has at least two primary heteroclinic orbits close to  $N$ .

As a second example, we consider the completely doubled case. We recall that the manifolds  $W^-$  and  $W^+$  are called *doubled* if  $W^- \setminus O^- = W^+ \setminus O^+ =: N$  as sets in  $M$ . They are called *completely doubled* if they are doubled and, in addition, the topologies in  $N$  induced from  $M$ ,  $W^+$ , and  $W^-$  coincide. This is equivalent to ask that the invariant manifolds have the same tangent spaces in  $N$ , that is,

$$T_m W^- = T_m W^+ \quad \forall m \in N.$$

In the completely doubled case, the finiteness condition holds by definition. Moreover,  $N$  is an  $n$ -dimensional manifold, and the invariant manifolds  $W^-$  and  $W^+$  have a clean intersection along  $N$  automatically. Since,  $N_i \simeq \mathbb{R} \times \mathbb{S}^{n-1}$  then  $\overline{N}_i \simeq \mathbb{S}^n$ , so the compactness condition holds and all the hypotheses of theorem 4. Concerning the quotient  $N/f$ , we note that if  $\det Df(O^\pm) > 0$ , the map  $f^s$  acts on  $N_i$  as  $(t, r) \rightarrow (t+1, r)$ , whereas if  $\det Df(O^\pm) < 0$ , it acts as  $(t, r) \rightarrow (t+1, \sigma(r))$ , where  $\sigma : \mathbb{S}^{n-1} \rightarrow \mathbb{S}^{n-1}$  is an orientation reversing involution. Hence,  $N/f = N_i/f^s \simeq \mathbb{S}^1 \times \mathbb{S}^{n-1}$  or its factor.

Theorem 4 is in the spirit of many analogous results, going back to Poincaré, on the perturbation of a manifold of periodic or homoclinic orbits. For the case of periodic orbits, see [23]. The case of homoclinic orbits of Hamiltonian systems and symplectic maps are contained, respectively, in [1] and [10]. The framework of exact symplectic

maps was studied recently by Xia [24], although his proof only covers the completely doubled case.

The proof of theorem 4 is given in the next section. There are two possible approaches: symplectic geometry (more elementary) and variational methods (more natural from the physical point of view). We have chosen the second one.

#### 4. Proof of the persistence for twist maps

First we define a variational problem for finding heteroclinic orbits of a twist map  $f$  from one hyperbolic periodic orbit  $O^-$  to another  $O^+$ . Here we do not assume that the assumptions of Theorem 4 are satisfied. What follows will be applied later to the maps  $f_\epsilon$  for  $\epsilon \geq 0$ .

Fix some sequence  $(y_i \in \tilde{Q})_{i \in \mathbb{Z}}$  such that  $y_i = x_i^-$  for large  $i < 0$ , and  $y_i = x_i^+$  for large  $i > 0$ . Fix a Riemannian metric  $\|\cdot\|^2 = \langle \cdot, \cdot \rangle$  on  $Q$ . Consider the function space  $X$  of sequences  $x \in \tilde{Q}^{\mathbb{Z}}$  such that

$$\sum_{i \in \mathbb{Z}} d^2(x_i, y_i) < \infty.$$

Then  $X$  is a Hilbert manifold. The tangent space at  $x \in X$  is

$$T_x X = \left\{ (\xi_i \in T_{x_i} \tilde{Q})_{i \in \mathbb{Z}} : \sum \|\xi_i\|^2 < \infty \right\}.$$

Define the scalar product  $\langle \xi, \eta \rangle$  of vectors  $\xi, \eta \in T_x X$  by  $\langle \xi, \eta \rangle = \sum \langle \xi_i, \eta_i \rangle$ . Then  $T_x X$  is a Hilbert space. A local chart for  $X$  with center  $x$  can be defined by the exponential map  $\phi : B \subset T_x X \rightarrow X$ ,  $\phi_i(\xi_i) = \exp_{x_i} \xi_i$ , where  $B$  is a small ball in  $T_x X$ .

If  $\tilde{Q} = \mathbb{R}^n$ , then  $X = l_2^n$  is a Hilbert space. For simplicity one can keep in mind this case. In general  $X$  is a Hilbert manifold modelled on  $l_2^n$ .

Suppose that the generating function  $L$  is normalized and we define the normalized action functional  $S$  on  $X$  by (16). We need to check that the sum is well-defined. This follows from a rearrangement of the series  $S(x)$ :

$$S(x) = \sum_{i \in \mathbb{Z}} \left( (D_2 L(y_{i-1}, y_i) + D_1 L(y_i, y_{i+1}))(x_i - y_i) + O_2(x_i - y_i) \right)$$

and the fact that  $y_i$  is a trajectory of the twist map for large  $|i|$ . Hence,  $S(x)$  is an absolutely convergent series for all  $x \in X$ .

**Lemma 4.**  $S \in C^2(X)$ , and its derivative is given by

$$S'(x)(\xi) = \sum_{i \in \mathbb{Z}} (D_2 L(x_{i-1}, x_i) + D_1 L(x_i, x_{i+1})) \xi_i, \quad \xi \in T_x X.$$

*Remark 3.* Hence,  $S'(x) = (D_2 L(x_{i-1}, x_i) + D_1 L(x_i, x_{i+1}))_{i \in \mathbb{Z}} \in T_x^* X$ . Note that  $S'(x)_i \in T_{x_i}^* \tilde{Q}$ , since the derivatives  $D_1 L, D_2 L$  lie in the cotangent space to  $\tilde{Q}$ . A similar formula holds for the gradient  $\nabla S(x) \in T_x X$ . Then instead of  $D_1 L$  one should write  $\nabla_1 L$ , where the gradient is taken with respect to the Riemannian metric on  $\tilde{Q}$ .

*Proof.* To check that  $S$  is differentiable, we calculate

$$S(\exp_x \xi) - S(x) = \sum_{i \in \mathbb{Z}} (D_2 L(x_{i-1}, x_i) + D_1 L(x_i, x_{i+1})) \xi_i + \sum_{i \in \mathbb{Z}} O_2(\xi_{i-1}, \xi_i).$$

Continuity of  $S' : X \rightarrow T^*X$  is evident. A similar computation shows that  $S \in C^2(X)$ .  $\square$

The following characterization is a direct consequence of Lemma 4.

**Lemma 5.** *Critical points of  $S : X \rightarrow \mathbb{R}$  are heteroclinic trajectories from  $O^-$  to  $O^+$ .*

If  $x \in X$  is a critical point for  $S$ , then the second derivative  $F = S''(x)$  is a correctly defined linear operator  $F : T_x X \rightarrow T_x^* X$  given by

$$(Fv)_i = D_{21}L(x_{i-1}, x_i)v_{i-1} + D_{12}L(x_i, x_{i+1})v_{i+1} + (D_{22}L(x_{i-1}, x_i) + D_{11}L(x_i, x_{i+1}))v_i. \quad (19)$$

We will need the following general result.

**Lemma 6.** *Let  $x$  be a critical point for  $S : X \rightarrow \mathbb{R}$ . Then the operator  $F = S''(x) : T_x X \rightarrow T_x^* X$  is a Fredholm operator, i.e.  $\dim \ker F < \infty$  and  $\dim T_x^* X / F(T_x X) < \infty$ .*

*Proof.* The operator  $F : T_x X \rightarrow T_x^* X$  has the form (19), namely

$$(Fv)_i = A_i^* v_{i-1} + A_{i+1} v_{i+1} + B_i v_i$$

where  $B_i = D_{22}L(x_{i-1}, x_i) + D_{11}L(x_i, x_{i+1})$  and  $A_i = D_{12}L(x_{i-1}, x_i)$ .

Hence,  $F$  is an elliptic difference operator provided that the twist condition  $\det A_i \neq 0$  holds along the trajectory  $(x_i)_{i \in \mathbb{Z}}$ . The first property of Fredholm operators follows from the twist condition: since the operators  $A_i$  are invertible, if  $F(v) = 0$ , then  $v$  is completely determined by  $v_{i-1}$  and  $v_i$ . Hence  $\dim \ker F \leq 2n$ . So only the second property of the Fredholm operators needs a proof. We will show that  $F$  is a sum of an invertible and a compact operator.

First suppose that  $x$  is a transverse heteroclinic trajectory. Then we will show that  $F$  is an isomorphism and  $\ker F = 0$ .

Let  $E_i^\pm = T_{p_i} W_i^\pm$ ,  $p_i = (x_{i-1}, x_i)$ . From now on we identify  $U$  with  $M$  and assume that  $W_i^\pm \subset U$ . By the transversality assumption,  $E_i^+ \cap E_i^- = \{0\}$ . Let us solve the equation  $Fv = w$  for given  $w \in T_x^* X$ .

First suppose that  $w_i = 0$  for  $i \neq j$ , i.e.  $w_i = w_i \delta_{ij}$ . Then for large  $|i|$  equation  $Fv = w$  means that  $v_i$  satisfies the variational equation for the heteroclinic trajectory  $x$ . Since  $v_i \rightarrow 0$  as  $|i| \rightarrow \infty$ , necessarily  $(v_j, v_{j+1}) \in E_{j+1}^+$  and  $(v_{j-1}, v_j) \in E_j^-$ . Then equation  $Fv = w$  implies

$$A_j^* v_{j-1} + A_{j+1} v_{j+1} + B_j v_j = w_j \quad (v_j, v_{j+1}) \in E_{j+1}^+ \quad (v_{j-1}, v_j) \in E_j^-.$$

Since  $E_j^+ \cap E_j^- = \{0\}$  by the transversality assumption, if all  $w_j = 0$ , then the only solution of these equations is  $v_j = v_{j-1} = v_{j+1} = 0$ . Hence for  $w_i = \delta_{ij} w_j$ , there exists a unique solution  $v_i = G_{ij} w_j$  where the Green function  $G_{ij}$  satisfies that  $\|G_{ij}\| \leq C e^{-\lambda|i-j|}$ ,

because trajectories in  $E_i^\pm$  tend to 0 exponentially as  $|i| \rightarrow \infty$ . Now for any  $w = (w_i)_{i \in \mathbb{Z}}$ , we get formally

$$v_i = \sum_{j \in \mathbb{Z}} G_{ij} w_j. \quad (20)$$

If  $\sum_{i \in \mathbb{Z}} \|w_i\|^2 < \infty$ , then  $\sum_{i \in \mathbb{Z}} \|v_i\|^2 < \infty$ , and so  $v \in T_x X$  given by (20) satisfies  $Fv = w$ . Thus  $F$  is invertible under the transversality assumption.

Now consider the general case. By a perturbation of  $L$  near the point  $p_j = (x_{j-1}, x_j)$  without changing the heteroclinic trajectory  $x$ , one can make the heteroclinic transverse:  $E_j^+ \cap E_j^- = \{0\}$ . This perturbation changes only a finite number of the operators  $A_i, B_i$ . Hence  $F$  is a sum of an invertible and a  $2n$ -dimensional operator. Thus  $F = S''(x)$  is a Fredholm operator.  $\square$

Let  $T : X \rightarrow X$  be the translation  $(x_i) \mapsto (x_{i+s})$ . Note that the translation  $(x_i) \mapsto (x_{i+1})$  does not take  $X$  to itself. Then  $T$  defines a discrete  $\mathbb{Z}$  group action on  $X$ . The functional  $S$  is  $T$ -invariant, so it is well-defined on the quotient space  $X/T$ . We do not discuss the topology of  $X/T$ , since now we restrict  $S$  to a compact submanifold of  $X$ .

Suppose that  $S \in C^2(X)$  has a  $T$ -invariant finite dimensional manifold  $Z \subset X$  of critical points. Thus  $Z$  consists of heteroclinic trajectories. Then for any  $x \in Z$ ,  $T_x Z \subset \ker S''(x)$ . Recall that a manifold  $Z$  of critical points of a function  $S$  is called a *non-degenerate critical manifold* [6, 17] if for any  $x \in Z$ , the operator  $F = S''(x) : T_x X \rightarrow T_x^* X$  has a closed range and  $T_x Z = \ker F$ . Then  $\dim T_x X / F(X) = \dim \ker F < \infty$ , so that  $F$  is a Fredholm operator. By Lemma 6 for any critical point  $x \in X$ ,  $S''(x)$  is a Fredholm operator, and so only the condition  $\ker F = T_x Z$  is non-trivial.

Recall that we identify  $M$  with  $U$ . Let

$$N_i = \{(x_{i-1}, x_i) \in U \subset (\tilde{Q} \times \tilde{Q})/G : x \in Z\} \subset W_i^+ \cap W_i^-$$

be the set of heteroclinic points in  $W_i^- \cap W_i^+$  corresponding to heteroclinic orbits in  $Z$ . There is a natural projection  $\pi_i : Z \rightarrow N_i$  given by  $\pi_i(x) = (x_{i-1}, x_i)$ . Note that  $N_{i+s} = N_i$  and  $\pi_i \circ T = f^s \circ \pi_i$ . The inverse map  $\phi_i : N_i \rightarrow Z \subset X$  is well-defined provided that the twist condition holds on  $N = \cup N_i$ . Obviously,  $Z/T = N_i/f^s$ .

**Lemma 7.** *Under the conditions of Theorem 4, the projection  $\pi_i : Z \rightarrow N_i$  is a  $C^1$  diffeomorphism.*

*Proof.* The map  $\pi_i$  is obviously  $C^1$ . Continuity of the map  $\phi_i = \pi_i^{-1} : N_i \rightarrow Z$  follows from the hyperbolicity of  $O^\pm$  and the finiteness assumption: if  $p = (x_{i-1}, x_i), \tilde{p} = (\tilde{x}_{i-1}, \tilde{x}_i) \in N_i$ , then the corresponding sequences  $(x_j)_{j \in \mathbb{Z}^+}, (\tilde{x}_j)_{j \in \mathbb{Z}^+}$  are exponentially close:  $\text{dist}(x_j, \tilde{x}_j) \leq C e^{-\lambda|j|} \text{dist}(p, \tilde{p})$ . Indeed, for large  $i \geq j$  we have  $(x_{i-1}, x_i) \in W_{i,\delta}^+$  and for  $i \leq -j$  we have  $(x_{i-1}, x_i) \in W_{i,\delta}^-$ . This implies that  $\phi_i \in C^0(N_i, Z)$ . Similarly, one can prove that  $\phi_i \in C^1(N_i, Z)$ .  $\square$



**Lemma 8.**  *$Z$  is a non-degenerate critical manifold for  $S$  if and only if  $W_i^-$  and  $W_i^+$  have a clean intersection along  $N_i$ , that is,*

$$T_m W_i^+ \cap T_m W_i^- = T_p N_i \quad \forall m \in N_i.$$

*If  $N_i \cup \{p_i^-, p_i^+\}$  is compact and the finiteness condition from Theorem 4 holds, then  $Z/T = N_i/f^s$  is a compact manifold.*

*Proof.* Note that  $\ker S''(x)$  is the set of solutions  $\xi = (\xi_i)_{i \in \mathbb{Z}}$  of the variational equation for the orbit  $x$  that tend to zero as  $i \rightarrow \pm\infty$ . But this holds if and only if the vector  $(\xi_{i-1}, \xi_i) \in T_{p_i}(\tilde{Q} \times \tilde{Q})$ ,  $p_i = (x_{i-1}, x_i)$ , belongs to  $T_{p_i} W^+ \cap T_{p_i} W^-$ . Hence

$$\dim \ker S''(x) = \dim(T_{p_i} W^+ \cap T_{p_i} W^-) = \dim N$$

provided that the intersection is clean.

The space  $N_i/f^s = N/f$  is a manifold since  $f^s$  defines a free discrete action of the group  $\mathbb{Z}$  on  $N_i$ . This is evident in the immersed topology from  $W_i^\pm$ , but it coincides with the embedded topology due to the finiteness condition. Compactness of  $N_i/f^s$  follows from the finiteness condition and the contraction property of  $f^s$  on  $W_{i,\delta}^\pm$ .

We note that if  $D = W_{i,\delta}^+ \setminus f^s(W_{i,\delta}^+)$  is a fundamental domain of  $W_i^+$ , then  $T^\mathbb{Z}(D \cap N_i) = N_i$ . Since  $\overline{D}$  is compact, the quotient space is also compact.  $\square$

Now let us consider the perturbed map  $f_\epsilon$ . Without loss of generality it will be assumed that the periodic orbit  $O^\pm$  does not change under the perturbation: for the perturbed trajectory,  $x_i^\pm(\epsilon) = x_i^\pm$ .

Indeed, suppose for simplicity that  $x_i \neq x_j$  for  $i \neq j \bmod s$ . Let  $\psi_\epsilon : Q \rightarrow Q$  be an isotopy such that  $\psi_\epsilon(x_i^\pm) = x_i^\pm(\epsilon)$  and let  $\theta_\epsilon = \psi_\epsilon \times \psi_\epsilon$ . Then the twist map  $\theta_\epsilon^{-1} \circ f_\epsilon \circ \theta_\epsilon$  with the generating function  $L \circ \theta_\epsilon$  has hyperbolic periodic trajectories  $x^\pm = (x_i^\pm)$  independent of  $\epsilon$ . With this reduction, the space  $X$  of sequences is independent of  $\epsilon$ .

The condition  $S_\epsilon(O_\epsilon^-) = S_\epsilon(O_\epsilon^+)$  implies that we have well-defined normalized action functionals  $S, S_\epsilon \in C^2(X)$  corresponding to  $L$  and  $L_\epsilon$  respectively, and

$$S_\epsilon = S + \epsilon S_1 + O(\epsilon^2)$$

is a  $C^2$ -small perturbation of  $S$ . The functional  $S_\epsilon$  is  $T$ -invariant on  $X$ . Hence  $S_\epsilon$  is defined on  $X/T$ . Now we use the following well-known result [6, 17], which follows from the Lyapunov–Schmidt reduction.

**Lemma 9.** *Let  $K$  be a compact non-degenerate critical manifold for a  $C^2$  function  $S$  on a Hilbert manifold. Let  $S_\epsilon = S + \epsilon S_1 + O(\epsilon^2)$  be a  $C^2$ -small perturbation of  $S$ . Then there exists a neighborhood  $U$  of  $K$  and a family of  $C^1$  manifolds  $K_\epsilon \subset U$ ,  $K_0 = K$ , given by a  $C^1$  embedding  $\phi_\epsilon : K \rightarrow U$ ,  $\phi_0 = \text{Id}_K$ , such that the critical points of  $S_\epsilon$  in  $U$  belong to  $K_\epsilon$  and are critical points of  $S_\epsilon \circ \phi_\epsilon : K \rightarrow \mathbb{R}$ .*

On the one hand, Lemma 9 obviously implies Theorem 4 if one puts  $K = Z/T$ .

On the other hand, we note that to any non-degenerate critical point of  $S_1|_K$  there corresponds a non-degenerate critical point of  $S_\epsilon$ , because

$$S_\epsilon \circ \phi_\epsilon = \text{constant} + \epsilon S_1|_K + O(\epsilon^2).$$

Therefore, if the generating function  $L_\epsilon$  of  $f_\epsilon$  has the form (17), then  $S_1 = P \circ \pi_i$ , where  $P$  is the Poincaré function (18) on  $N_i$  and  $\pi_i : Z \rightarrow N_i$  is the projection. This implies Theorem 3.

## 5. Proofs of the results on billiards

This section contains the proofs of the results on perturbed elliptic billiards presented in section 2. The main theorem on billiards (theorem 2) follows directly from the lemmas 1 and 2 and the main theorem on twist maps (theorem 4). To prove the lemmas, we need a description of the homoclinic set in two very particular cases: planar billiards inside non-circular ellipses and billiards inside prolate ellipsoids. This has to do with the fact that the set  $N$  defined in (11) is the union of the homoclinic sets of several planar and prolate sub-billiards. Then the proof of lemma 1 becomes trivial. Next, we shall generalize a result of Devaney [12] to obtain lemma 2.

### 5.1. Proof of theorem 2

The phase space  $M_\epsilon$  of the perturbed billiard map  $f_\epsilon$  depends on  $\epsilon$ , but making a symplectic transformation we may assume that  $f_\epsilon : M \rightarrow M$ . By lemmas 1 and 2, the set  $N_j$  verifies the hypotheses of theorem 4. Recall the Ljusternik-Schnirelmann categories

$$\text{cat}(N_j/f) = \text{cat}(\mathbb{S}^1 \times \mathbb{S}^{s_j-1}) = \begin{cases} 4 & \text{if } s_j = 1 \\ 3 & \text{if } s_j > 1 \end{cases}.$$

This ends the proof of theorem 2. It remains to prove lemmas 1 and 2.

### 5.2. The homoclinic set of planar elliptic billiards

The results in this subsection are very old. They can be found in the books [13, 21], so we skip the proofs. The style in our presentation may look a little pedantic, but it is the best suitable for the extension contained in the next subsection.

Our goal is to describe the homoclinic set  $H = \text{Hom}_f(P)$  of the two-periodic set  $P = \{m_+, m_-\}$  of the billiard map  $f$  inside the non-circular ellipse

$$Q = \left\{ q = (x, y) \in \mathbb{R}^2 : \frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1 \right\} \quad \alpha > \beta > 0. \quad (21)$$

Let  $e = (1 - \beta^2/\alpha^2)^{1/2}$  and  $\gamma = \sqrt{\alpha^2 - \beta^2}$  be the eccentricity and the semi-focal distance of this ellipse. Let  $\lambda = (1+e)(1-e)^{-1}$  be the characteristic multiplier of  $P$ . Let  $h = \ln \lambda$  be the characteristic exponent of  $P$ .

We are dealing with a planar billiard, so we do not need sub-indices in the coordinates. We write the points of the billiard phase space (3) as  $m = (q, p)$  with

$q = (x, y) \in Q$  and  $p = (u, v) \in \mathbb{S}^1 \subset \mathbb{R}^2$ . The elliptic planar billiard has only one first integral functionally independent. For instance,

$$F(m) = v^2 - \gamma^{-2}(xv - yu)^2$$

is a good choice (compare with (7)). It turns out that  $H = F^{-1}(0) \setminus P$ .

For visualization purposes, we identify the billiard phase space with the annulus

$$\mathbb{A} = \{(\varphi, \rho) \in \mathbb{T} \times \mathbb{R} : \rho^2 < \gamma^2 \sin^2 \varphi + \beta^2\}$$

by means of the relations  $q = q(\varphi) := (\alpha \cos \varphi, \beta \sin \varphi)$  and  $\rho = \langle \dot{q}(\varphi), p \rangle = |\dot{q}(\varphi)| \cos \vartheta$ , where  $\vartheta \in (0, \pi)$  is the angle between the tangent vector  $\dot{q}(\varphi)$  and the unitary velocity  $p$ . In these coordinates, the above first integral is

$$F(\varphi, \rho) = \sin^2 \varphi - \gamma^{-2} \rho^2.$$

The phase portrait of the planar elliptic billiard map, considered as a diffeomorphism on the annulus, is displayed in figure 1. It shows that the homoclinic set

$$H = F^{-1}(0) \setminus P = \{(\varphi, \rho) \in \mathbb{A} : \rho = \pm \gamma \sin \varphi \neq 0\}$$

has four connected components which are  $f^2$ -invariant, but not  $f$ -invariant. The arrows in the figure show the  $f^2$ -dynamics on  $H$ . Thus,  $H$  splits into two disjoint heteroclinic sets:  $H = H_- \cup H_+$ , where

$$H_- = \text{Het}_{f^2}(m_+, m_-) \quad H_+ = \text{Het}_{f^2}(m_-, m_+). \quad (22)$$

(Given two fixed points  $a_-$  and  $a_+$  of an invertible map  $\phi : M \rightarrow M$ , we denote  $\text{Het}_\phi(a_-, a_+) = \{m \in M : \lim_{k \rightarrow \pm\infty} \phi^k(m) = a_\pm\}$ .)

The following lemma is a corollary of the above comments (see also figure 1). We recall that  $\mathbb{S}^0 = \{-1, +1\} \subset \mathbb{R}$ .

**Lemma 10.** *The homoclinic set  $H = H_- \cup H_+$  of the billiard map  $f$  inside any non-circular ellipse verifies the following properties:*

(i) *It contains four curves (a couple in  $H_-$  and another couple in  $H_+$ ) and*

$$H/f \simeq H_-/f^2 \simeq H_+/f^2 \simeq \mathbb{S}^1 \times \mathbb{S}^0.$$

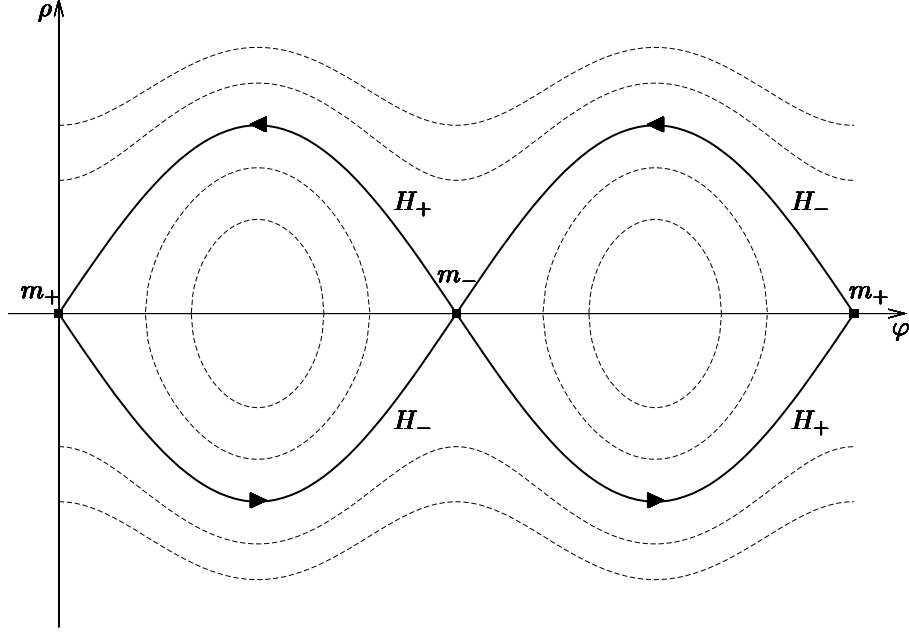
(ii)  *$H \cup P$  is compact. In fact,  $H_\pm \cup P \simeq \mathbb{S}^1$ .*

(iii) *Given any neighborhood  $U$  of  $P$ , there exists  $k > 0$  such that  $f^k(H \setminus U) \subset U$  for all integer  $|k| > k_0$ .*

Finally, we need an explicit expression for the billiard dynamics on the homoclinic set. The following formulae can be found in many papers. See, for instance, [9].

Let  $q : \mathbb{R} \times \mathbb{S}^0 \rightarrow Q$  and  $p : \mathbb{R} \times \mathbb{S}^0 \rightarrow \mathbb{S}^1$  be the maps

$$q(t, \sigma) = (x(t), \sigma y(t)) \quad p(t, \sigma) = (u(t), \sigma v(t))$$



**Figure 1.** The phase portrait of the planar elliptic billiard map  $f : \mathbb{A} \rightarrow \mathbb{A}$  for  $\alpha = 1$  and  $\beta = 0.8$ . The solid squares denote the hyperbolic fixed points  $m_{\pm}$ . The thick lines denote the heteroclinic connections  $H = H_- \cup H_+$ . The solid arrows denote the dynamics of the map on the connections.

where  $x(t) = \alpha \tanh t$ ,  $y(t) = \beta \operatorname{sech} t$ ,  $u(t) = \tanh(t - h/2)$ , and  $v(t) = \operatorname{sech}(t - h/2)$ . Then, the maps  $m_{\pm} := \pm(q, p) : \mathbb{R} \times \mathbb{S}^0 \rightarrow M$  are natural parameterizations of the heteroclinic sets  $H_{\pm}$ , i.e.,  $m_{\pm} : \mathbb{R} \times \mathbb{S}^0 \rightarrow H_{\pm}$  are analytic diffeomorphisms such that

$$f(m_{\pm}(t, \sigma)) = m_{\mp}(t + h, \sigma).$$

For further reference, we also need to compute the tangent spaces to the stable and unstable invariant curves at the hyperbolic periodic points. These tangent spaces are one-dimensional. Thus, it suffices to find some vectors  $\dot{m}^{\pm} \in T_P W^{\pm}$ . To begin with, let us express the previous natural parameterizations in terms of the variable  $r = e^t$ . We also recall that  $\lambda = e^h$ . Let  $\bar{q} : \mathbb{R} \rightarrow Q$  and  $\bar{p} : \mathbb{R} \rightarrow \mathbb{S}^1$  be the maps

$$\bar{q}(r) = \left( \alpha \frac{1 - r^2}{1 + r^2}, \frac{2\beta r}{1 + r^2} \right) \quad \bar{p}(r) = \left( \frac{\lambda - r^2}{\lambda + r^2}, \frac{2\lambda^{1/2}r}{\lambda + r^2} \right).$$

Let  $\bar{m} = (\bar{q}, \bar{p}) : \mathbb{R} \rightarrow M$ . Then the diffeomorphisms  $m_{\pm}^{u,s} : \mathbb{R} \rightarrow W^{u,s}(m_{\pm})$  defined by  $m_{\pm}^u(r) = \pm \bar{m}(r)$  and  $m_{\pm}^s(r) = \mp \bar{m}(1/r)$  verify that

$$m_{\pm}^{s,u}(0) = m_{\pm} \quad f(m_{\pm}^s(r)) = m_{\mp}^s(r/\lambda) \quad f(m_{\pm}^u(r)) = m_{\mp}^u(\lambda r)$$

so we can take  $\dot{m}^+ = \frac{1}{2}\lambda^{-1/2}\dot{m}_+^s(0)$  and  $\dot{m}^- = \frac{1}{2}\lambda^{1/2}\dot{m}_+^u(0)$ . It is trivial to see that

$$\dot{m}^{\pm} = (\dot{q}^{\pm}, \dot{p}^{\pm}) = ((0, \eta^{\pm}), (0, 1)) \quad \eta^{\pm} := \lambda^{\pm 1/2}\beta. \quad (23)$$

### 5.3. The homoclinic set of prolate elliptic billiards

Here we describe the homoclinic set  $H = \text{Hom}_f(P)$  of the two-periodic hyperbolic set  $P = \{m_+, m_-\}$  of the billiard map  $f$  inside the (non-spherical) prolate ellipsoid

$$Q = \left\{ q \in \mathbb{R}^{n+1} : \frac{q_0^2}{\alpha^2} + \frac{q_1^2 + \dots + q_n^2}{\beta^2} = 1 \right\} \quad \alpha > \beta > 0. \quad (24)$$

The main idea is that, since this prolate ellipsoid is, in some rough sense, the ellipse (21) times the sphere  $\mathbb{S}^{n-1}$ , then its corresponding homoclinic set will be, also in some rough sense, the one corresponding to the ellipse times the sphere  $\mathbb{S}^{n-1}$ .

Given any unit vector  $\sigma \in \mathbb{S}^{n-1} \subset \mathbb{R}^n$ , we consider the plane

$$\Pi_\sigma = \{q = (x, y \cdot \sigma) : x, y \in \mathbb{R}\} \subset \mathbb{R}^{n+1}$$

and the section  $Q_\sigma = Q \cap \Pi_\sigma$ . The plane  $\Pi_\sigma$  contains the diameter of the prolate ellipsoid, whereas the section  $Q_\sigma$  is an ellipse whose semi-axis have lengths  $\alpha$  and  $\beta$ , and whose foci are  $(\pm\gamma, 0, \dots, 0)$ , where  $\gamma = \sqrt{\alpha^2 - \beta^2}$  is the semi-focal distance.

If two consecutive impact points are on  $Q_\sigma$ , the others impact points also are. This observation is the key to relate the billiard dynamics on the homoclinic set corresponding to the prolate ellipsoid (24) with the billiard dynamics on the homoclinic set corresponding to the ellipse (21), which has been given in the previous subsection.

Concretely, if  $q : \mathbb{R} \times \mathbb{S}^{n-1} \rightarrow Q$  and  $p : \mathbb{R} \times \mathbb{S}^{n-1} \rightarrow \mathbb{S}^n$  are the maps

$$q(t, \sigma) = (x(t), y(t) \cdot \sigma) \quad p(t, \sigma) = (u(t), v(t) \cdot \sigma)$$

where  $x(t) = \alpha \tanh t$ ,  $y(t) = \beta \text{sech } t$ ,  $u(t) = \tanh(t - h/2)$ , and  $v(t) = \text{sech}(t - h/2)$ , then the maps  $m_\pm = \pm(q, p) : \mathbb{R} \times \mathbb{S}^{n-1} \rightarrow M$  are natural parameterizations of the heteroclinic sets  $H_\pm$ . That is,  $m_\pm = \pm(q, p) : \mathbb{R} \times \mathbb{S}^{n-1} \rightarrow H_\pm$  are analytic diffeomorphisms such that

$$f(m_\pm(t, \sigma)) = m_\mp(t + h, \sigma).$$

Besides, the limits  $\lim_{t \rightarrow -\infty} m_\pm(t, \sigma) = m_\mp$  and  $\lim_{t \rightarrow +\infty} m_\pm(t, \sigma) = m_\pm$  are uniform in  $\sigma \in \mathbb{S}^{n-1}$ . Hence, the generalization of lemma 10 to the case of prolate ellipsoids reads as follows.

**Lemma 11.** *The homoclinic set  $H = H_- \cup H_+$  of the billiard map  $f$  inside any (non-spheric) prolate ellipsoid verify the following properties:*

(i) *It contains two  $n$ -dimensional connected submanifolds:  $H_-$  and  $H_+$ . Besides,*

$$H/f \simeq H_-/f^2 \simeq H_+/f^2 \simeq \mathbb{S}^1 \times \mathbb{S}^{n-1}.$$

(ii)  *$H \cup P$  is compact. In fact,  $H_\pm \cup P \simeq \mathbb{S}^n$ .*

(iii) *Given any neighbourhood  $U$  of  $P$ , there exists  $k_0 > 0$  such that  $f^k(H \setminus U) \subset U$  for all integer  $|k| > k_0$ .*

To end this subsection, we mention that the only difference between the ellipses and the prolate ellipsoids is that  $\mathbb{S}^0$  has two disconnected points, whereas  $\mathbb{S}^{n-1}$  is connected for  $n > 1$ . Due to this, the homoclinic set contains four curves (loops) in the first case, and two  $n$ -dimensional connected submanifolds in the second one.

#### 5.4. Proof of lemma 1

Lemma 1 follows directly, due to its local character, from lemma 10 on the homoclinic set of billiards inside non-circular ellipses, and from lemma 11 on the homoclinic set of billiards inside prolate ellipsoids.

#### 5.5. Proof of lemma 2

The phase space  $M$  contains points  $m = (q, p) \in \mathbb{R}^{2n+2}$  such that  $q = (q_0, \dots, q_n) \in Q$  and  $p = (p_0, \dots, p_n) \in \mathbb{S}^n$ . That is  $\parallel$ ,

$$M = \left\{ m = (q, p) \in \mathbb{R}^{2n+2} : \sum_{i=0}^n d_i^{-2} q_i^2 = 1 \quad \sum_{i=0}^n p_i^2 = 1 \right\}.$$

Thus, tangent vectors to the phase space can also be considered as elements of  $\mathbb{R}^{2n+2}$ . They will be denoted with a dot:

$$\dot{m} = (\dot{q}, \dot{p}) = ((\dot{q}_0, \dots, \dot{q}_n), (\dot{p}_0, \dots, \dot{p}_n)) \in T_m M \subset \mathbb{R}^{2n+2}.$$

Finally, if  $q = (q_0, \dots, q_n) \in Q$  and  $p = (p_0, \dots, p_n) \in \mathbb{S}^n$ , we shall use the notation  $\hat{q}_j = (q_{a_j}, \dots, q_{b_j}) \in \mathbb{R}^{s_j}$  and  $\hat{p}_j = (p_{a_j}, \dots, p_{b_j}) \in \mathbb{R}^{s_j}$ , for  $j = 1, \dots, l$ .

The invariant manifolds  $W^-$  and  $W^+$  are contained in the zero level sets of the first integrals, that is,  $W^\pm \subset \cap_{j'=1}^l Z_{j'}$ . We are going to investigate the structure of  $Z_{j'}$  at points  $m \in N_j$ . The cases  $j' = j$  and  $j' \neq j$  are very different.

**Lemma 12.** *The zero level set  $Z_j$  is a smooth  $(2n - s_j)$ -dimensional submanifold of the phase space  $M$  at any point  $m \in N_j$ . Besides, the intersection  $N_j = (Z_j \setminus P) \cap \Pi_j$  is transverse in  $M$ . In particular,  $T_m N_j = T_m Z_j \cap T_m \Pi_j$  for all  $m \in N_j$ .*

*Proof.* We distinguish two cases:  $s_j = 1$  and  $s_j > 1$ .

*Case  $s_j = 1$ .* Then  $\# \llbracket a_j, b_j \rrbracket = 1$ . Let  $i$  be the integer such that  $\llbracket a_j, b_j \rrbracket = \{i\}$ . The invariant section  $\Pi_j \subset M$ , the zero level set  $Z_j = F_i^{-1}(0)$ , and the intersection  $N_j = (Z_j \setminus P) \cap \Pi_j$  can be written as

$$\begin{aligned} \Pi_j &= \left\{ m \in \mathbb{R}^{2n+2} : \begin{array}{l} d_0^{-2} q_0^2 + d_i^{-2} q_i^2 = 1 \quad p_0^2 + p_i^2 = 1 \\ q_{i'} = p_{i'} = 0 \text{ for all } i' \neq 0, i \end{array} \right\} \\ Z_j &= \left\{ m \in \mathbb{R}^{2n+2} : \begin{array}{l} \sum_{i'} d_{i'}^{-2} q_{i'}^2 = 1 \quad \sum_{i'} p_{i'}^2 = 1 \\ p_i^2 + \sum_{i' \neq i} \frac{(q_i p_{i'} - q_{i'} p_i)^2}{d_i^2 - d_{i'}^2} = 0 \end{array} \right\} \end{aligned} \quad (25)$$

$\parallel$  In fact,  $M$  is just one connected component of the set defined by those equations—namely, the component of points  $m = (q, p)$  such that  $p$  is directed outward  $Q$  at  $q$ —, but we shall skip this detail for the sake of notation.

$$N_j = \left\{ m \in \mathbb{R}^{2n+2} : \begin{array}{l} d_0^{-2} q_0^2 + d_i^{-2} q_i^2 = 1 \quad p_0^2 + p_i^2 = 1 \\ p_i^2 = \frac{(q_i p_0 - q_0 p_i)^2}{d_0^2 - d_i^2} \neq 0 \\ q_{i'} = p_{i'} = 0 \text{ for all } i' \neq 0, i \end{array} \right\}. \quad (26)$$

Once we have these equations, the claims of the lemma are mere computations.

*Case  $s_j > 1$ .* Let  $I = \llbracket a_j, b_j \rrbracket$ . Then the set  $A_j$  is formed by the points in which the angular momenta  $K_{(i,i')}(m) = q_i p_{i'} - q_{i'} p_i$  vanish for all  $i, i' \in I$ . The key idea is to realize that if  $m = (q, p) \in A_j$ , the vectors  $\hat{q}_j, \hat{p}_j \in \mathbb{R}^{s_j}$  are linearly dependent: there exist  $(\tilde{q}_j, \tilde{p}_j) \in \mathbb{R}^2$  and  $\sigma_j \in \mathbb{S}^{s_j-1}$  such that  $\hat{q}_j = \tilde{q}_j \sigma_j$  and  $\hat{p}_j = \tilde{p}_j \sigma_j$ . Besides,  $A_j$  is a smooth submanifold of  $M$  at any point  $m$  such that  $(\tilde{q}_j, \tilde{p}_j) \neq (0, 0)$ .

From now on, the points  $(\tilde{q}_j, \tilde{p}_j) \in \mathbb{R}^2$  and  $\sigma_j \in \mathbb{S}^{s_j-1}$  have always this meaning. In particular,  $\tilde{q}_j^2 = \sum_{i \in I} q_i^2$ ,  $\tilde{p}_j^2 = \sum_{i \in I} p_i^2$ , and  $S_j(m) = \tilde{p}_j^2 + \sum_{i' \notin I} \frac{(\tilde{q}_j p_{i'} - q_{i'} \tilde{p}_j)^2}{\tilde{d}_j^2 - d_{i'}^2}$  for all  $m \in A_j$ . Hence, we can write the invariant section  $\Pi_j$ , the zero level set  $Z_j = S_j^{-1}(0) \cap A_j$  and the intersection  $N_j = (Z_j \setminus P) \cap \Pi_j$  as follows:

$$\begin{aligned} \Pi_j &= \left\{ m \in \mathbb{R}^{2n+2} : \begin{array}{l} d_0^{-2} q_0^2 + \tilde{d}_j^{-2} \sum_{i \in I} q_i^2 = 1 \\ p_0^2 + \sum_{i \in I} p_i^2 = 1 \\ q_{i'} = p_{i'} = 0 \text{ for all } i' \notin I \cup \{0\} \end{array} \right\} \\ Z_j &= \left\{ m \in \mathbb{R}^{2n+2} : \begin{array}{l} \hat{q}_j = \tilde{q}_j \sigma_j \text{ and } \hat{p}_j = \tilde{p}_j \sigma_j \\ \sum_{i'} d_{i'}^{-2} q_{i'}^2 = \tilde{d}_j^{-2} \tilde{q}_j^2 + \sum_{i' \notin I} d_{i'}^{-2} q_{i'}^2 = 1 \\ \sum_{i'} p_{i'}^2 = \tilde{p}_j^2 + \sum_{i' \notin I} p_{i'}^2 = 1 \\ \tilde{p}_j^2 + \sum_{i' \notin I} \frac{(\tilde{q}_j p_{i'} - q_{i'} \tilde{p}_j)^2}{\tilde{d}_j^2 - d_{i'}^2} = 0 \end{array} \right\} \\ N_j &= \left\{ m \in \mathbb{R}^{2n+2} : \begin{array}{l} \hat{q}_j = \tilde{q}_j \sigma_j \text{ and } \hat{p}_j = \tilde{p}_j \sigma_j \\ d_0^{-2} q_0^2 + \tilde{d}_j^{-2} \tilde{q}_j^2 = 1 \\ p_0^2 + \tilde{p}_j^2 = 1 \\ \tilde{p}_j^2 = \frac{(\tilde{q}_j p_0 - q_0 \tilde{p}_j)^2}{d_0^2 - \tilde{d}_j^2} \neq 0 \\ q_{i'} = p_{i'} = 0 \text{ for all } i' \notin I \cup \{0\} \end{array} \right\} \end{aligned} \quad (27)$$

From these expressions, it is again straightforward to check that the lemma holds.  $\square$

When  $j' \neq j$  the structure of  $Z_{j'}$  at points  $m \in N_j$  is more involved. It turns out that  $Z_{j'}$  consists, in a neighbourhood of  $N_j$ , of two smooth submanifolds  $Z_{j'}^\pm$  of codimension  $s_{j'}$  in the phase space  $M$ . Besides, these submanifolds have a transverse intersection along

$$\begin{aligned} \Lambda_{j'} &= \{m \in M : q_{i'} = p_{i'} = 0 \text{ for all } i' \in \llbracket a_{j'}, b_{j'} \rrbracket\} \\ &= \{m \in M : \hat{q}_{j'} = \hat{p}_{j'} = 0\} \end{aligned}$$

Finally, the invariant manifold  $W^\pm$  is a submanifold of  $Z_{j'}^\pm$  and  $\Pi_j = \cap_{j' \neq j} \Lambda_{j'}$ . (This result was proved by Devaney [12] when  $Q$  is a generic ellipsoid.) Roughly speaking, these are the main steps in the proof of the following result.

**Lemma 13.**  $T_m W^- \cap T_m W^+ \subset \cap_{j' \neq j} T_m \Lambda_{j'} = T_m \Pi_j$  for all  $m \in N_j$ .

*Proof.* We must prove that  $T_m W^- \cap T_m W^+ \subset T_m \Lambda_{j'}$  for all  $m \in N_j$  and  $j' \neq j$ .

As in the previous lemma, we distinguish two cases:  $s_{j'} = 1$  and  $s_{j'} > 1$ .

*Case  $s_{j'} = 1$ .* Then  $\# \llbracket a_{j'}, b_{j'} \rrbracket = 1$ . Let  $i'$  be the integer such that  $\llbracket a_{j'}, b_{j'} \rrbracket = \{i'\}$ . Thus  $Z_{j'} = F_{i'}^{-1}(0)$  and  $\Lambda_{j'} = \{m \in M : q_{i'} = p_{i'} = 0\}$ . Let  $I = \llbracket a_j, b_j \rrbracket$ . (The case  $s_j = 1$ —or equivalently,  $I = \{i\}$ —is not excluded. In that case,  $\tilde{q}_j = q_i$  and  $\tilde{p}_j = p_i$ .)

To begin with, we shall work on the Euclidean space  $\mathbb{R}^{2n+2}$ . For instance, the first integral  $F_{i'} : M \rightarrow \mathbb{R}$  can be extended to this Euclidean space, because it is polynomial in the coordinates  $m = (q, p)$ . We denote this extension by  $\bar{F}_{i'}$ . We also consider the  $2n$ -dimensional linear subspace

$$\bar{\Lambda}_{j'} = \{m \in \mathbb{R}^{2n+2} : q_{i'} = p_{i'} = 0\}$$

Clearly,  $\Lambda_{j'} = \bar{\Lambda}_{j'} \cap M$  and  $T_m \Lambda_{j'} = T_m \bar{\Lambda}_{j'} \cap T_m M$  for all  $m \in \Lambda_{j'}$ .

The function  $\bar{F}_{i'}$  vanishes together with its first partial derivatives along  $\bar{\Lambda}_{j'}$ . So  $\bar{\Lambda}_{j'}$  is a critical manifold for  $\bar{F}_{i'}$ . We are going to show that  $\bar{\Lambda}_{j'}$  is a non-degenerate critical manifold in the sense of Bott.

Let  $\bar{\pi}_{j'} : \mathbb{R}^{2n+2} \rightarrow \bar{\Lambda}_{j'}$  be the canonical projection along the  $(q_{i'}, p_{i'})$ -plane. If  $m \in \bar{\Lambda}_{j'}$ , let  $\bar{P}_{j'}^m = \bar{\pi}_{j'}^{-1}(m)$  be the orthogonal plane to  $\bar{\Lambda}_{j'}$  at  $m$ . We are going to study the intersections of the zero set  $\bar{Z}_{j'}$  with the bi-dimensional slices  $\bar{P}_{j'}^m$  for  $m \in \bar{\Lambda}_{j'}$ . Let  $\psi_{j'}^m$  be the restriction of  $\bar{F}_{i'}$  to  $\bar{P}_{j'}^m$ . Then the point  $m$  is a critical point of  $\psi_{j'}^m$  and, if  $m$  is close enough to  $N_j$ , it is a saddle point:  $\det[d^2 \psi_{j'}^m(m)] < 0$ .

We compute this determinant using the coordinates  $(q_{i'}, p_{i'})$ , namely

$$\Delta_{j'}(m) := -\det[d^2 \psi_{j'}^m(m)] = -\det \begin{pmatrix} \alpha_{j'}(m) & \beta_{j'}(m) \\ \beta_{j'}(m) & \gamma_{j'}(m) \end{pmatrix}$$

where  $\alpha_{j'}(m) = \frac{\partial^2 \bar{F}_{i'}}{\partial q_{i'}^2}(m)$ ,  $\beta_{j'}(m) = \frac{\partial^2 \bar{F}_{i'}}{\partial q_{i'} \partial p_{i'}}(m)$  and  $\gamma_{j'}(m) = \frac{\partial^2 \bar{F}_{i'}}{\partial p_{i'}^2}(m)$ .

Using that  $N_j \subset \Pi_j \cap A_j$ , we find that

$$\begin{aligned} \alpha_{j'}(m) &= \sum_{i \neq i'} \frac{2p_i^2}{d_{i'}^2 - d_i^2} = \frac{2p_0^2}{d_{i'}^2 - d_0^2} + \frac{2\tilde{p}_j^2}{d_{i'}^2 - \tilde{d}_j^2} \\ \beta_{j'}(m) &= -\sum_{i \neq i'} \frac{2q_i p_i}{d_{i'}^2 - d_i^2} = -\frac{2q_0 p_0}{d_{i'}^2 - d_0^2} - \frac{2\tilde{q}_j \tilde{p}_j}{d_{i'}^2 - \tilde{d}_j^2} \\ \gamma_{j'}(m) &= 2 + \sum_{i \neq i'} \frac{2q_i^2}{d_{i'}^2 - d_i^2} = 2 + \frac{2q_0^2}{d_{i'}^2 - d_0^2} + \frac{2\tilde{q}_j^2}{d_{i'}^2 - \tilde{d}_j^2}. \end{aligned}$$

On the other hand, the points  $m \in N_j$  verify the equations  $(d_0^2 - \tilde{d}_j^2)\tilde{p}_j^2 = (\tilde{q}_j p_0 - q_0 \tilde{p}_j)^2$  and  $p_0^2 + \tilde{p}_j^2 = 1$ , see (27). Therefore,

$$\begin{aligned} \Delta_{j'}(m) &= \beta_{j'}^2(m) - \alpha_{j'}(m)\gamma_{j'}(m) \\ &= 4 \left( \frac{p_0^2}{d_0^2 - d_{i'}^2} + \frac{\tilde{p}_j^2}{\tilde{d}_j^2 - d_{i'}^2} + \frac{2q_0 \tilde{q}_j p_0 \tilde{p}_j - q_0^2 \tilde{p}_j^2 - \tilde{q}_j^2 p_0^2}{(d_{i'}^2 - d_0^2)(d_{i'}^2 - \tilde{d}_j^2)} \right) \\ &= 4 \left( \frac{1 - \tilde{p}_j^2}{d_0^2 - d_{i'}^2} + \frac{\tilde{p}_j^2}{\tilde{d}_j^2 - d_{i'}^2} - \frac{(q_0 \tilde{p}_j - \tilde{q}_j p_0)^2}{(d_{i'}^2 - d_0^2)(d_{i'}^2 - \tilde{d}_j^2)} \right) \end{aligned}$$



$$\begin{aligned}
&= \frac{4}{d_0^2 - d_{i'}^2} + 4 \left( \frac{(d_0^2 - \tilde{d}_j^2) \tilde{p}_j^2}{(d_0^2 - d_{i'}^2)(\tilde{d}_j^2 - d_{i'}^2)} - \frac{(d_0^2 - \tilde{d}_j^2) \tilde{p}_j^2}{(d_0^2 - d_{i'}^2)(\tilde{d}_j^2 - d_{i'}^2)} \right) \\
&= \frac{4}{d_0^2 - d_{i'}^2} > 0
\end{aligned}$$

for all  $m \in N_j$ . This implies that there exists a neighborhood  $\bar{V}$  of  $N_j$  in  $\bar{\Lambda}_{j'}$  such that  $m$  is a saddle point of  $\psi_{j'}^m$ , for any  $m \in \bar{V}$ . In particular, given any  $m \in \bar{V}$  there exists a neighborhood  $\bar{U}_m$  of  $m$  in  $\bar{P}_{j'}^m$  such that the set  $(\psi_{j'}^m)^{-1}(0) \cap \bar{U}_m$  contains two smooth curves  $C_{j',m}^\pm$  which have a transverse crossing at  $m$ . The tangent lines to these curves at  $m$  verify the linear equation  $\dot{q}_{i'} = \mu_{j'}^\pm(m) \dot{p}_{i'}$ , where

$$\mu_{j'}^\pm(m) = \frac{-\beta_{j'}(m) \pm \sqrt{\Delta_{j'}(m)}}{\alpha_{j'}(m)}$$

are the roots of  $\alpha_{j'}(m)\mu^2 + 2\beta_{j'}(m)\mu + \gamma_{j'}(m) = 0$ . It is important to remark that these roots are real and different, because  $\Delta_{j'}(m) > 0$ .

Then  $\bar{U} = \cup_{m \in \bar{V}} \bar{U}_m$  is a neighborhood of  $N_j$  in  $\mathbb{R}^{2n+2}$  and  $\bar{Z}_{j'}^\pm = \cup_{m \in \bar{V}} C_{j',m}^\pm$  are a couple of hypersurfaces of  $\mathbb{R}^{2n+2}$  such that  $\bar{F}_{i'}^{-1}(0) \cap \bar{U} = \bar{Z}_{j'}^- \cup \bar{Z}_{j'}^+$  and the intersection  $\bar{Z}_{j'}^- \cap \bar{Z}_{j'}^+ = \cup_{m \in \bar{V}} (C_{j',m}^- \cap C_{j',m}^+) = \bar{V} = \bar{U} \cap \bar{\Lambda}_{j'}$  is transverse. In fact,

$$T_m \bar{Z}_{j'}^\pm = \{ \dot{m} \in \mathbb{R}^{2n+2} : \dot{q}_{i'} = \mu_{j'}^\pm(m) \dot{p}_{i'} \} \quad \forall m \in N_j.$$

Since  $M$  is transverse to  $\bar{\Lambda}_{j'}$  in  $\mathbb{R}^{2n+2}$ , it follows that  $M$  is also transverse to  $\bar{Z}_{j'}^\pm$  in  $\mathbb{R}^{2n+2}$ . Thus the set  $Z_{j'} = F_{i'}^{-1}(0) = \bar{F}_{i'}^{-1}(0) \cap M$  has the promised structure in the neighborhood  $U = \bar{U} \cap M$  of  $N_j$  in  $M$ . It suffices to take  $Z_{j'}^\pm = \bar{Z}_{j'}^\pm \cap M$ , and so

$$T_m Z_{j'}^\pm = \{ \dot{m} \in T_m M : \dot{q}_{i'} = \mu_{j'}^\pm(m) \dot{p}_{i'} \} \quad \forall m \in N_j. \quad (28)$$

Once we have shown this structure, it becomes clear that the tangent spaces  $T_m W^-$  and  $T_m W^+$  are contained in the union of the tangent spaces  $T_m Z_{j'}^-$  and  $T_m Z_{j'}^+$  for all  $m \in N_j$ , because the invariant manifolds  $W^-$  and  $W^+$  are submanifolds of  $M$  contained in the zero level set  $Z_{j'}$ . Let us assume that we have performed this construction in such a way that the matching up of signs is the expected one:

$$T_m W^\pm \subset T_m Z_{j'}^\pm \quad \forall m \in N_j \quad \forall j' \neq j. \quad (29)$$

Under that assumption, and using that the roots  $\mu_{j'}^-(m)$  and  $\mu_{j'}^+(m)$  never coincide, we get that  $T_m W^- \cap T_m W^+ \subset T_m Z_{j'}^- \cap T_m Z_{j'}^+ = \{ \dot{m} \in T_m M : \dot{q}_{i'} = \dot{p}_{i'} = 0 \} = T_m \Lambda_{j'}$ , for all  $m \in N_j$  and for all  $j' \neq j$ . This finishes the proof of the lemma when  $s_{j'} = 1$ .

Hence, it remains to prove that (29) holds. To see this, by continuity, it suffices to check that it holds for the points in the hyperbolic periodic set  $P$ , since  $P \subset \bar{N}_j$ .

Firstly, let us compute the value of the roots  $\mu_{j'}^\pm$  at  $P$ . If  $m = m_-$  or  $m = m_+$ , then  $\alpha_{j'}(m) = -2/(d_0^2 - d_{i'}^2)$  and  $\beta_{j'}(m) = 2d_0/(d_0^2 - d_{i'}^2)$ . So  $\mu_{j'}^\pm(m) = d_0(1 \mp e_{i'})$ , where  $e_{i'} = (1 - d_{i'}^2/d_0^2)^{1/2}$ . In particular,

$$\mu_{j'}^+(m) \cdot \mu_{j'}^-(m) = d_{i'}^2 \quad \mu_{j'}^+(m)/\mu_{j'}^-(m) = 1/\lambda_{i'}.$$

This implies that  $\mu_{j'}^\pm(m_+) = \mu_{j'}^\pm(m_-) = \lambda_{i'}^{\mp 1/2} d_{i'}$  and so

$$T_P Z_{j'}^\pm = \{\dot{m} \in T_P M : \dot{q}_{i'} = \eta_{i'}^\pm \dot{p}_{i'}\} \quad \eta_{i'}^\pm = \lambda_{i'}^{\mp 1/2} d_{i'}. \quad (30)$$

Secondly, by comparison with the planar case, it is easy to find that

$$\dot{m} = ((0, \dots, 0, \overset{i'}{\eta_{i'}^\pm}, 0, \dots, 0), (0, \dots, 0, \overset{i'}{1}, 0, \dots, 0)) \in T_P W^\pm \quad (31)$$

using that the set  $\Pi_{j'}$  is invariant by  $f$ . In fact, the sub-billiard  $f|_{\Pi_{j'}}$  is identical to the planar billiard map inside the ellipse

$$Q_{j'} = \left\{ q \in \mathbb{R}^{n+1} : \frac{q_0^2}{d_0^2} + \frac{q_{i'}^2}{d_{i'}^2} = 1, \quad q_k = 0 \text{ for } k \neq 0, i' \right\}.$$

The eccentricity of the ellipse  $Q_{j'}$  is  $e_{i'} = (1 - d_{i'}^2/d_0^2)^{1/2}$  and the characteristic multiplier of the planar map  $f|_{\Pi_{j'}}$  is  $\lambda_{i'} = (1 + e_{i'})(1 - e_{i'})^{-1}$ . Hence, (31) follows from the results contained in subsection 5.2. Concretely, compare with formula (23).

Finally, the matching up (29) is obtained from the combination of (30) and (31).

*Case  $s_{j'} > 1$ .* Then  $Z_{j'} = S_{j'}^{-1}(0) \cap A_{j'}$  and  $\Lambda_{j'} = \{m \in M : \hat{q}_{j'} = \hat{p}_{j'} = 0\}$ . Let  $I' = \llbracket a_{j'}, b_{j'} \rrbracket$  and  $I = \llbracket a_j, b_j \rrbracket$ . (As before, the possibility  $s_j = 1$  is included.)

We shall use the same method than before, although we are going to consider slides of dimension  $2s_{j'}$ , instead of bi-dimensional ones. We shall also work on the Euclidean space  $\mathbb{R}^{2n+2}$ . Thus, let  $\bar{S}_{j'} : \mathbb{R}^{2n+2} \rightarrow \mathbb{R}$  be the natural extension of the first integral  $S_{j'} : M \rightarrow \mathbb{R}$  and let  $\bar{Z}_{j'} = \bar{S}_{j'}^{-1}(0) \cap \bar{A}_{j'}$ , where

$$\bar{A}_{j'} = \{(q, p) \in \mathbb{R}^{2n+2} : q_i p_{i'} = q_{i'} p_i \text{ for all } i, i' \in I'\}.$$

Given any point  $m = (q, p) \in \bar{A}_{j'}$ , there exist  $(\tilde{q}_{j'}, \tilde{p}_{j'}) \in \mathbb{R}^2$  and  $\sigma_{j'} \in \mathbb{S}^{s_{j'}-1}$  such that  $\hat{q}_{j'} = \tilde{q}_{j'} \sigma_{j'}$  and  $\hat{p}_{j'} = \tilde{p}_{j'} \sigma_{j'}$ . We also consider the  $2n$ -dimensional linear subspace

$$\bar{\Lambda}_{j'} = \{m \in \mathbb{R}^{2n+2} : \hat{q}_{j'} = \hat{p}_{j'} = 0\}.$$

Let  $\bar{\pi}_{j'} : \mathbb{R}^{2n+2} \rightarrow \bar{\Lambda}_{j'}$  be the canonical projection along the  $(\hat{q}_{j'}, \hat{p}_{j'})$ -coordinates. If  $m \in \bar{\Lambda}_{j'}$ , let  $\bar{P}_{j'}^m = \bar{\pi}_{j'}^{-1}(m)$  be the  $2s_{j'}$ -dimensional orthogonal slice to  $\bar{\Lambda}_{j'}$  at  $m$ . Then

$$(\hat{q}_{j'}, \hat{p}_{j'}) \in \bar{Z}_{j'} \cap \bar{P}_{j'}^m \iff \begin{cases} \hat{q}_{j'} = \tilde{q}_{j'} \sigma_{j'} \\ \hat{p}_{j'} = \tilde{p}_{j'} \sigma_{j'} \\ \tilde{\psi}_{j'}^m(\tilde{q}_{j'}, \tilde{p}_{j'}) = 0 \end{cases}$$

where the function  $\tilde{\psi}_{j'}^m : \mathbb{R}^2 \rightarrow \mathbb{R}$  is defined by

$$\tilde{\psi}_{j'}^m(\tilde{q}_{j'}, \tilde{p}_{j'}) = \tilde{p}_{j'}^2 + \sum_{i \notin I'} \frac{(\tilde{q}_{j'} p_i - q_i \tilde{p}_{j'})^2}{\tilde{d}_{j'}^2 - d_i^2}.$$

The origin is a critical point of  $\tilde{\psi}_{j'}^m$ . Moreover, a computation very similar to the one performed in the previous case shows that

$$\det[d^2 \tilde{\psi}_{j'}^m(0, 0)] = -4/(d_0^2 - \tilde{d}_{j'}^2) < 0 \quad \forall m \in N_j.$$

This implies that there exists a neighborhood  $\bar{V}$  of  $N_j$  in  $\bar{\Lambda}_{j'}$  such that the origin is a saddle point of  $\tilde{\psi}_{j'}^m$ , for any  $m \in \bar{V}$ . In particular, given any  $m \in \bar{V}$  there exists

a neighborhood  $\tilde{U}_m$  of the origin in  $\mathbb{R}^2$  such that the set  $(\tilde{\psi}_{j'}^m)^{-1}(0) \cap \tilde{U}_m$  contains two smooth curves  $\tilde{C}_{j',m}^\pm$  which have a transverse crossing at the origin. Since  $\tilde{\psi}_{j'}^m$  is even, there exists a couple of smooth functions  $\phi_{j',m}^\pm$  such that

$$\tilde{C}_{j',m}^\pm = \{(\tilde{q}_{j'}, \tilde{p}_{j'}) \in \mathbb{R}^2 : \tilde{p}_{j'} = \tilde{q}_{j'} \phi_{j',m}^\pm(\tilde{q}_{j'}^2)\}.$$

Therefore, there exists a neighborhood  $\bar{U}_m$  of  $m$  in  $\bar{P}_{j'}^m$  such that the set  $\bar{Z}_{j'} \cap \bar{U}_m$  contains the two smooth  $s_{j'}$ -dimensional submanifolds

$$C_{j',m}^\pm = \{(\hat{q}_{j'}, \hat{p}_{j'}) \in \mathbb{R}^{2s_{j'}} : \hat{p}_{j'} = \phi_{j',m}^\pm(|\hat{q}_{j'}|^2) \hat{q}_{j'}\}$$

which have a transverse intersection at the point  $m$ . The tangent spaces to these submanifolds are defined by the linear equations

$$\dot{q}_{i'} = \mu_{j',m}^\pm(m) \dot{p}_{i'} \quad \forall i' \in I'$$

where the quantities  $\mu_{j',m}^-(m) = \phi_{j',m}^-(0)$  and  $\mu_{j',m}^+(m) = \phi_{j',m}^+(0)$  never coincide. (The explicit expressions for  $\mu_{j',m}^\pm(m)$  are almost equal to the ones obtained in the previous case; it suffices to change  $d_{i'}$  by  $\tilde{d}_{j'}$ .)

Then  $\bar{U} = \cup_{m \in \bar{V}} \bar{U}_m$  is a neighborhood of  $N_j$  in  $\mathbb{R}^{2n+2}$  and  $\bar{Z}_{j'}^\pm = \cup_{m \in \bar{V}} C_{j',m}^\pm$  are a couple of manifolds of codimension  $s_{j'}$  in  $\mathbb{R}^{2n+2}$  such that  $\bar{Z}_{j'} \cap \bar{U} = \bar{Z}_{j'}^- \cup \bar{Z}_{j'}^+$  and the intersection  $\bar{Z}_{j'}^- \cap \bar{Z}_{j'}^+ = \cup_{m \in \bar{V}} (C_{j',m}^- \cap C_{j',m}^+) = \bar{V} = \bar{U} \cap \bar{\Lambda}_{j'}$  is transverse:

$$T_m \bar{Z}_{j'}^\pm = \{\dot{m} \in \mathbb{R}^{2n+2} : \dot{q}_{i'} = \mu_{j',m}^\pm(m) \dot{p}_{i'} \text{ for all } i' \in I'\} \quad \forall m \in N_j.$$

In particular, the set  $Z_{j'} = S_{j'}^{-1}(0) \cap A_{j'} = \bar{Z}_{j'} \cap M$  has the promised structure in the neighborhood  $U = \bar{U} \cap M$  of  $N_j$  in  $M$ . It suffices to take  $Z_{j'}^\pm = \bar{Z}_{j'}^\pm \cap M$ , and so

$$T_m Z_{j'}^\pm = \{\dot{m} \in T_m M : \dot{q}_{i'} = \mu_{j',m}^\pm(m) \dot{p}_{i'} \text{ for all } i' \in I'\} \quad \forall m \in N_j.$$

The end of the proof follows the lines given in the previous case. Firstly, it turns out that the tangent spaces of the invariant manifolds  $W^-$  and  $W^+$  are contained in the tangent spaces of  $Z_{j'}^-$  and  $Z_{j'}^+$ , respectively. That is,

$$T_m W^\pm \subset T_m Z_{j'}^\pm \quad \forall m \in N_j \quad \forall j' \neq j.$$

Under that assumption, and since the quantities  $\mu_{j',m}^-(m)$  and  $\mu_{j',m}^+(m)$  never coincide, we get that  $T_m W^- \cap T_m W^+ \subset T_m Z_{j'}^- \cap T_m Z_{j'}^+ = \{\dot{m} \in T_m M : \dot{q}_{i'} = \dot{p}_{i'} = 0\} = T_m \Lambda_{j'}$ , for all  $m \in N_j$  and for all  $j' \neq j$ . This finishes the proof of the lemma when  $s_{j'} > 1$ .  $\square$

The proof of lemma 2 ends with the following corollary.

**Corollary 2.**  $T_m W^- \cap T_m W^+ = T_m N_j$  for all  $m \in N_j$ .

*Proof.* Let  $m$  be any point in  $N_j$ . It suffices to prove that

$$T_m N_j \subset T_m W^- \cap T_m W^+ \subset T_m Z_j \cap T_m \Pi_j = T_m N_j$$

because then, the above inclusions are, in fact, equalities.

Firstly, inclusion  $T_m W^- \cap T_m W^+ \subset T_m \Pi_j$  was obtained in lemma 13. Next, inclusions  $T_m N_j \subset T_m W^- \cap T_m W^+ \subset T_m Z_j$  follow from  $N_j \subset W^\pm \subset Z_j$ . Finally, equality  $T_m Z_j \cap T_m \Pi_j = T_m N_j$  was obtained in lemma 12.  $\square$

## Acknowledgments

This work has been partially supported by the INTAS grant 00-221. One of the authors (SB) has also been partially supported by the RFBR grant 02-01-00400 and NSF grant 0300319. Two of the authors (AD and RR-R) have also been partially supported by the Spanish grant DGICYT BFM2000-0805 and the Catalan grant CIRIT 2001SGR-70.

## References

- [1] Ambrosetti A and Badiale M 1998 Homoclinics: Poincaré-Melnikov type results via a variational approach *Ann. Inst. H. Poincaré Anal. Non Linéaire* **15** 233–252
- [2] Birkhoff G D 1927 *Dynamical Systems (Am. Math. Soc. Coll. Pub. vol 9)* (Providence, RI: American Mathematical Society)
- [3] Bolotin S 2000 Infinite number of homoclinic orbits to hyperbolic invariant tori of Hamiltonian systems *Regul. Chaotic Dyn.* **5** 139–56
- [4] Bolotin S 2001 Symbolic dynamics near minimal hyperbolic invariant tori of Lagrangian systems *Nonlinearity* **14** 1123–40
- [5] Bolotin S, Delshams A, Fedorov Yu and Ramírez-Ros R 2002 Bi-asymptotic billiard orbits inside perturbed ellipsoids *Progress in Nonlinear Science (vol I)* ed L Lerman and L Shilnikov (Nizhny Novgorod) pp 48–62
- [6] Bott R 1954 Non-degenerate critical manifolds *Ann. of Math.* **60** 248–61
- [7] Burns K and Weiss H 1995 A geometric criterion for positive topological entropy *Comm. Math. Phys.* **172** 95–118
- [8] Delshams A, Fedorov Yu and Ramírez-Ros R 2001 Homoclinic billiard orbits inside symmetrically perturbed ellipsoids *Nonlinearity* **14** 1141–95
- [9] Delshams A and Ramírez-Ros R 1996 Poincaré-Melnikov-Arnold method for analytic planar maps *Nonlinearity* **9** 1–26
- [10] Delshams A and Ramírez-Ros R 1997 Melnikov potential for exact symplectic maps *Comm. Math. Phys.* **190** 213–45
- [11] Delshams A and Ramírez-Ros R 1998 Homoclinic orbits for twist maps and billiards *Symmetry and Perturbation Theory '96 (Quaderni G.N.F.M. vol 54)* ed D Bambusi and G Gaeta (Florence, Consiglio Nazionale delle Ricerche) pp 46–86
- [12] Devaney R L 1978 Transversal homoclinic orbits in an integrable system *Amer. J. Math.* **100** 631–42
- [13] Kozlov V and Treshchëv D 1991 *Billiards: a Genetic Introduction to the Dynamics of Systems with Impacts (Transl. Math. Monographs vol 89)* (Providence, RI: American Mathematical Society)
- [14] Kozlova T V 1998 Nonintegrability of a rotating elliptical billiard *J. Appl. Math. Mech.* **62** 81–5
- [15] Levallois P 1997 Calcul d’une fonction de Melnikov et de ses zéros pour une perturbation algébrique du billard elliptique *Ergodic Theory Dynam. Systems* **17** 435–44
- [16] Lomelí H 1996 Perturbations of elliptic billiards *Physica D* **99** 59–80
- [17] Meyer K 1967 Kritische Mannigfaltigkeiten in Hilbertmannigfaltigkeiten *Math. Ann.* **170** 45–66
- [18] Poincaré H 1890 Sur le problème des trois corps et les équations de la dynamique *Acta Math.* **13** 1–271
- [19] Robnik M and Berry M V 1985 Classical billiards in magnetic fields *J. Phys. A* **18** 1361–78
- [20] Smale S 1965 Diffeomorphisms with many periodic points *Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse)* ed S S Cairns (Princeton, NJ: Princeton University Press) pp 63–80
- [21] Tabachnikov S 1995 *Billiards* Panor. Synth. **1** vi+142
- [22] Tabanov M 1994 Splitting of separatrices for Birkhoff’s billiard in symmetric convex domain, closed to an ellipse *Chaos* **4** 595–606

- [23] Weinstein A 1978 Bifurcations and Hamilton's principle *Math. Z.* **159** 235–48
- [24] Xia Z 2000 Homoclinic points and intersections of Lagrangian submanifolds *Discrete Contin. Dynam. Systems* **6** 243–53